Exploring HIV Latency and Dissecting the Contribution of APOBEC3G to HIV Evolution Using Humanized Mice Models

---

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Audrey Fahrny

von

Le Locle, NE

Promotionskommission

Prof. Dr. Maries van den Broek (Vorsitz)
Prof. Dr. med. Roberto Speck (Leitung der Dissertation)
Prof. Dr. Michael O. Hottiger
Prof. Dr. Christian Münz

Zürich, 2018

Table of Contents

SUMMARY

Despite the advent of highly effective combined antiretroviral therapy (cART), HIV remains a major global health concern, for which we still do not have a cure. The major barrier to curing HIV is a long-lived reservoir of latently infected cells. Latently infected cells in patients are very rare ($1/10^6$ CD4+ T cells) and cannot be clearly distinguished from their uninfected counterparts as we lack a ubiquitous cell surface marker to identify them. Reservoir cells are impervious to host immune surveillance and cART, and most importantly, they can be induced to re-establish a systemic infection upon cART interruption. Thus, currently the gold standard in management of HIV is life-long cART, which besides being associated with important health, economic and social burdens, is available to only about half of HIV-infected individuals and millions of new infections occurring every year globally reflects reality. To halt the HIV pandemic, a cure is needed. To this end, one of the main approaches involves purging the latent HIV reservoir. However, progress in this field is hindered by the difficulty in studying HIV latency *ex vivo* and the lack of truly representative *in vitro* models. Another major approach to HIV cure is vaccination. However, HIV's propensity to rapidly diversify within its host confers it with an immense evolutionary potential that enables HIV to evade immune recognition and develop drug resistances, and has thus far thwarted efforts in vaccine development. Thus, HIV diversification within its host is another major barrier to HIV eradication. The work in this thesis focuses on two features of HIV that preclude its eradication, namely, HIV latency and HIV diversification.

Our first aim was to develop a HIV reporter humanized mouse model in which genuine latently infected cells can be identified, characterized and isolated. Such a model would provide valuable insights into the nature of the cells harboring latent HIV, as well as the mechanism that support latency, which are necessary for development of effective strategies to purge the latent reservoir. As humanized mice (hu-mice) provide a uniquely mutable system in which HIV infection can be studied in the context of a functional human immune system, we hypothesized that they would serve as a good basis for developing a HIV reporter model particularly suited to studying HIV latency. Thus, we introduced a Cre recombinase (Cre) inducible fluorescent transgene into the NSG humanized mouse model and used a recombinant HIV strain that encodes Cre for the infections. The concept behind this design is that when a cell carrying the inducible fluorescent transgene is infected with a HIV encoding Cre, the Cre subsequently present in the cell mediates recombination at loxP sites flanking the fluorescent reporter gene, resulting in the thereafter-constitutive expression of the fluorescent protein. Thus, in the model, infected cells are marked *in vivo* and can be isolated based on the expression of the fluorescent reporter gene. Then, by treating infected reporter hu-mice with cART, latently infected cells can be studied and isolated from various deep tissues. We demonstrated that our HIV reporter hu-mouse model is functional, as we could specifically recognize and isolate productively infected cells in our model by flow cytometry and fluorescence-activated cell sorting. However when used in a latency scenario, the

resolution of our model was not high enough to enable the unambiguous detection and purification of latently infected cells based on the expression of the reporter gene. Thus, further optimizations must be implemented to make our model more suitable for studying HIV latency.

Our second aim was to dissect the contribution of the host deaminase APOBEC3G (A3G) to HIV diversification. A3G is a component of the host innate anti-viral immune system and a potent restriction factor of HIV. A3G introduces extensive G-to-A mutations into the HIV genome, which generally render the virus defective. However, HIV effectively counteracts A3G with its accessory proteins Vif, which targets A3G for proteasomal degradation. Although Vif is indispensable for replication of HIV *in vivo*, Vif alleles with suboptimal A3G-neutralization abilities have been isolated from patients and are associated with HIV diversification and cART failure. Thus, we hypothesized that A3G-induced mutagenesis can enhance HIV diversification and HIV's ability to adapt to selective pressures. To this end, we infected hu-mice with isogenic HIV clones differing only in their ability to counteract A3G and characterized the phenotype and genotype of the different viruses either in natural infections or in the context of an antiretroviral drug (3TC) to impose a selective pressure on our system. We found that in the absence of a selective pressure, the HIV clone with only partial activity against A3G was attenuated over time compared to wild type HIV, however it displayed a superior 3TC-adaptation phenotype. We show that these phenotypes are supported by increased viral diversity characterized by an A3G-mutational footprint and the faster emergence of a 3TC-resistance mutation that results from a G-to-A mutation in a dinucleotide context favored by A3G. Our results support the notion that sublethal A3G activity can enhance the evolutionary potential of HIV *in vivo*.

ZUSAMMENFASSUNG

Die HIV Pandemie ist, trotz den heute verfügbaren hochaktiven antiretroviralen Therapien (ART), die sehr wirksam die HIV assoziierte Morbidität und Mortalität verringern, nicht unter Kontrolle und bleibt vor allen in Afrika ein immenses Problem für die dortige Bevölkerung. Diese ART vermögen zwar die HIV Vermehrung im Patienten zu unterdrücken, führen aber zu keiner Heilung. Eine der essentiellen Hürden für eine Heilung stellt das latente HIV Reservoir dar, d.h. HIV kann in bestimmten Zelltypen jahrelang überdauern. Zudem gibt es nur sehr wenig latent HIV infizierte Zellen in einem Patienten; Schätzwerte gehen von ca. einer latent HIV infizierten Zelle auf 10^6 CD4+ T-Lymphozyten aus. Diese latent infizierten Zellen sind von anderen nicht-infizierten Zellen nicht aktuell unterscheidbar, da kein Zelloberflächen Marker existiert, der diese latent HIV-infizierte Zellen eindeutig identifizieren könnte. Diese latent infizierten Zellen sind auch nicht für das Immunsystem (Abwehrsystem) nicht erkennbar. Wie oben erwähnt, sind auch die herkömmlichen antiretroviralen Therapien unwirksam, da diese rein die produktive HIV Vermehrung unterdrücken, aber keinen Effekt auf das latente HIV Reservoir haben. Dieses führt nach Absetzen der ART zu einer erneuten HIV Infektion und entsprechend zu Fortschreiten der HIV Erkrankung. Entsprechend muss ART von den HIV-infizierten Patienten lebenslang eingenommen werden. Diese Therapien sind eine enorme gesundheitliche und soziale Belastung für die Patienten sowie eine Last für die Volkswirtschaft eines Landes. Des Weiteren ist ART aktuell nur für ca. die Hälfte aller HIV-infizierte Menschen verfügbar und es infizieren sich immer noch jährlich Millionen von Menschen weltweit neu mit HIV. Eine Heilung ist daher notwendig um die HIV Pandemie zu besiegen. Seit längeren wird studiert wie man dieses latente HIV Reservoir angreifen kann. Der Wissensgewinn in diesem Gebiet ist jedoch limitiert, da das Studium der HIV Latenz *in vivo* aufgrund der Seltenheit und fehlenden Markern der latent HIV-infizierten Zellen sehr schwierig ist, und *in vitro* Modelle nur beschränkt die HIV Latenz simulieren können.

Unser erstes Ziel war es, ein Mausmodell zu generieren, das es ermöglicht, latent HIV-infizierte Zellen anhand eines Reporters zu identifizieren und diese dann zu charakterisieren und isolieren. Ein solches Modell wäre sehr wertvoll, um die die Latenz mechanistisch zu erforschen und ist zudem eine Notwendigkeit für die Entwicklung effizienter Therapien, die das latente HIV Reservoir zu eliminieren können. Das lymphatische Immunsystem in humanisierten Mäusen ist mittlerweile sehr gut manipulierbar und wir haben entsprechend humanisierte Mäuse eingesetzt, um ein solches Mausmodell für das Studium der HIV-1 Latenz zu entwickeln. Um dies zu erreichen haben wir ein Cre Recombinase (Cre) induzierbares Reportergen in die lymphatischen Zellen der humanisierten Mäuse eingebracht, welches nach Infektion mit einem rekombinanten HIV-Stamm, dass das Cre in seinem Genom enthält und in infizierten Zellen exprimiert, aufleuchtet. So ist dieses Reportergen flankiert von loxP Elementen, welche durch die Cre erkannt werden. Daraufhin kommt es zu einer sequenzspezifischen Rekombination des Reportergens und dieses kann dann in den infizierten Zellen exprimiert werden, in anderen Worten, nur Zellen die mit dem HIV infiziert

werden, leuchten auf. Der Promoter für die Expression des Reportgens ist konstitutiv aktiv und führt zu einer Transkription des Reportergens unabhängig ob es sich um eine produktive oder latente Infektion handelt. Die konstitutive Expression des Reportergens sollte dann auch die Isolation latent HIV infizierten Zellen erlauben, nachdem die virale Replikation unterdrückt ist. Somit wäre ein Mausmodell verfügbar, das mit ART behandelt werden kann, und in dem nur noch latent HIV-infizierte Zellen vorhanden sind. Wir konnten zeigen, dass das Mausmodell basierend auf diesem Reportersystem funktionierte und es uns möglich war mittels des Reportergens produktiv infizierte Zellen mittels Durchflusszytometrie zu identifizieren und zu isolieren. Leider war dieses Reportersystem nicht robust genug, um in humanisierten Mäuse, die mit ART behandelt wurden, eindeutig latent HIV-infizierten Zellen zu identifizieren und isolieren. Daher muss dieses System noch weiter für das *in vivo* Studium latent infizierter Zellen optimiert werden.

Ein zweites Ziel dieser Arbeit war das Studium der Deaminase APOBEC3G (A3G), das in Wirtszellen exprimiert wird, inwieweit diese zur Diversifikation von HIV beiträgt. A3G ist eine Komponente des angeborenen Immunsystems und ein potenter HIV Restriktionsfaktor. In akut infizierten Zellen führt A3G zu Guanin-zu-Adenosin (G-zu-A) Mutationen im HI-viralen Genom, welche zu replikationsdefekten HI-viralen Virionen führen, die sich nicht mehr vermehren können. HIV kann jedoch den Effekt von A3G mittels eines viralen, akzessorischen Proteins namens Vif kompensieren. Vif führt zum proteosomalen Abbau von A3G in der infizierten Zelle und ist daher essentiell für die virale Replikation. Es wurden aber auch Vif Varianten in HIV-1 Patienten identifiziert, die nur mässig A3G neutralisieren können und diese korrelierten waren mit einer verstärkten HIV Diversifikation und einem Therapieversagen. Bekanntlich führt eine erhöhte Diversifikation der HI-viralen Proteine eher zu einem Therapieversagen. Entsprechend war die Arbeitshypothese, das A3G-induzierte Mutagenese zu einer erhöhten Mutationsrate mit erhöhter Diversifikation führt und somit dem Selektionsdruck einer antiretroviralen Monotherapie standhält. Um dieser Frage nach zu gehen, haben wir humanisierte Mäuse mit isogenen HIV Stämmen infiziert, die sich nur in der Vif basierten Eigenschaft A3G zu neutralisieren, unterscheiden. Wir haben dann in diesen Mäusen die viralen Phäno- und Genotypen entweder ohne Therapie oder unter Therapie mit der antiretroviralen Substanz 3TC untersucht. 3TC wurde hier verabreicht, um, wie oben ausgeführt, den Selektionsdruck auf das Virus zu erhöhen. Wir konnten experimentell nachweisen, dass der HIV-Stamm mit einer partiellen Vif basierten anti-A3G Aktivität im Vergleich zu dem Wildtyp-Stamm in unbehandelten Mäusen, inhibiert war. Jedoch war die partielle Vif Aktivität vorteilhaft sobald die Mäuse mit 3TC behandelt wurden. So konnten wir zeigen, dass der HIV-Stamm mit einer partiellen Vif Aktivität zu einer A3G basierten, höheren Diversifikation der HI Viren führte und damit zu einem schnelleren 3TC-Therapieversagen, das auf die favorisierte A3G induzierte G-zu-A Mutation zurückzuführen war. Die Ergebnisse dieser Arbeit unterstützen die These, dass eine subletale A3G Aktivität die Diversität und damit das evolutionäre Potenzial von HIV *in vivo* fördert.

1. INTRODUCTION

1.1. The HIV pandemic and current treatment strategies

In 1981, a series of clinical case reports of *Pneumocystis carinii* pneumonia (PCP) and Kaposi sarcomas (KS) primarily in young and apparently healthy homosexual men and intravenous drug abusers living in New York and California, rose the suspicion of a new retrovirus causing severe immunodeficiency [1]. PCP and KS are rare diseases that occur otherwise only in severely immunosuppressed patients. By 1983, the human immunodeficiency virus (HIV) type 1 (HIV-1, hereon forth simply referred to as HIV) was discovered as the causative agent [2, 3] and has become a major global health concern, with a death toll of at least 35 million individuals so far [4]. HIV originated in Africa from numerous cross-species transmissions of simian immunodeficiency virus (SIV) strains from their natural hosts, African nonhuman primates, to humans [5]. According to the World Health Organization (WHO), in 2016 there were approximately 36.7 million HIV-infected individuals and 1.8 million new infections globally, with sub-Saharan African countries suffering the greatest HIV/AIDS burden [4].

HIV is a complex retrovirus that primarily infects CD4+ T lymphocytes and left untreated, it leads to a complex immune system dysfunction with chronic immune activation and inflammation, characterized by progressive depletion of the CD4+ T cells. This progressive immunodeficiency eventually culminates in the acquired immunodeficiency syndrome (AIDS) within 2 to 15 years in the majority of infected individuals [6]. AIDS is a severe state of immune system dysfunction, characterized by CD4+ T cell counts below 200 cells/ml of blood, high susceptibility to opportunistic infections and the development of AIDS-associated cancers and a fatal outcome in all patients.

The advent of antiretroviral drugs (ARVs) and development of combination antiretroviral therapy (cART) in 1996 were major medical breakthroughs for the management of HIV infection, drastically curbing HIV transmission, mortality and morbidity [7, 8]. cART effectively suppresses viral replication and currently is the only effective treatment against HIV-associated immunodeficiency, yet it cannot eradicate HIV. This is due to the rapid seeding of a latent HIV reservoir that persists despite prolonged viral suppression with cART [9-14], evidenced by the rapid reemergence of infectious virus upon treatment interruption [15, 16]. The major barrier to HIV eradication, and therefore HIV cure, is this latent reservoir. Without the complete elimination of the latent reservoir, cART must be taken lifelong, which is associated with a number of problems, such as adverse events, high costs, access to HIV clinics and care, social stigmatism, and emergence of drug-resistances in case of non-adherence [17, 18]. In addition, the WHO reports that in 2016 only 53%

of HIV-infected individuals received cART [19], owing mainly to the difficulty in provision and access to HIV-testing and care in Sub-Saharan Africa.

The other major barrier to HIV eradication is HIV's immense evolutionary potential, driven mostly by exceptionally high mutation and replication rates [20, 21] and recombination between co-packaged genomes [22, 23]. Indeed, although HIV transmission is usually established by a single founder virus, within a very short period of time the HIV strains in the newly infected host become genetically distinguishable from one another, resulting in a highly heterogeneous pool of viral quasispecies. In this context, HIV has an exceptional ability to escape immune recognition, to adapt to environmental and antiretroviral drug pressures, and renders universal vaccination strategies ineffective [24, 25]. Therefore, a greater understanding of HIV latency, as well as HIV's mechanisms of diversification will be crucial in developing effective novel eradication and vaccination strategies.

As current cART and vaccination strategies cannot eradicate HIV, a cure is urgently needed to halt the HIV pandemic. Currently three main strategies are being explored: i) vaccines against HIV (reviewed in [26]); ii) gene therapies aiming to either excise integrated provirus from host cells or render host cells resistant to HIV infection (reviewed in [27]); and iii) drugs to purge or fully block the latent reservoir using latency reversing or latency-inducing agents, respectively (reviewed in [28, 29]). The latter will be discussed in more detail in section 1.5.4. HIV cure strategies can be subdivided into two categories, based on the aspired end goal. Sterile cure approaches aim to eradicate all cells harboring replication-competent HIV, whilst functional cure strategies aim to achieve sustained viral control in the absence of cART by, for example, boosting and supplementing the host immune system.

1.2. HIV structure and genome

HIV is a human-tropic lentivirus, belonging to the *Retroviridae* family. Distinguishing features of retroviruses are their pseudodiploid single-stranded RNA genomes and the necessity for chromosomal integration of their viral genetic material for productive viral replication. The viral RNA genome is reverse transcribed into double-stranded DNA by a viral reverse transcriptase (RT) and a viral integrase (IN) then catalyzes its permanent integration into the host genome. Thereafter the proviral genome (i.e. provirus) serves as a template for the synthesis of viral mRNAs and proteins required for production of progeny virions, which is orchestrated by cellular and viral proteins. Besides the three essential polyproteins Gag, Pol and Env, which encoding structural gene products common to all retroviruses, the HIV genome also encodes six accessory proteins: Tat (transactivator of transcription), Rev (regulator of expression of viral proteins), Nef (negative factor), Vpr (viral protein r), Vif (viral infectivity factor), and Vpu (viral protein unique)

(Figure 1). Therefore, HIV is categorized as a complex retrovirus. The entire length of the HIV genome is approximately 9.7kb. Gag codes for structural proteins necessary to the assembly of the virion infrastructure: the matrix (MA), capsid (CA, i.e. p24), nucleocapsid (NC), and p6. Pol codes for enzymes essential to replication and virion maturation: the RT, IN and protease (PR). Env encodes the surface and transmembrane glycoproteins (gp), gp120 and gp41, respectively, that are part of the viral envelope and necessary for binding and entry into target cells. HIV's accessory proteins are multifunctional and essential to virus infectivity and propagation *in vivo* (reviewed in [30]). Briefly, regulatory proteins Tat and Rev are essential for viral replication and virion production, meditating efficient transcription of full-length viral mRNAs and nuclear export of unspliced and partially spliced viral mRNAs, respectively. Vpr is primarily responsible for the nuclear import of the viral complementary DNA (cDNA) (i.e. as part of the preintegration complex (PIC)), causes G2 phase cell cycle arrest and enables viral replication in non-dividing cells. Vpu and Nef both modulate, via distinct pathways, cell surface levels of the CD4 molecule, tetherin and MHC-class I, thereby preventing superinfection and antibody-dependent cell-mediated cytotoxicity (ADCC), as well as promoting virion release and supporting immune escape. Importantly, Nef also enhances viral infectivity by counteracting the host restriction factor SERINC5 [31] and facilitates viral assembly. Vif is primarily responsible for counteracting a set of host restriction factors, the APOBEC3 deaminases, which are part of the human antiviral innate immunity.

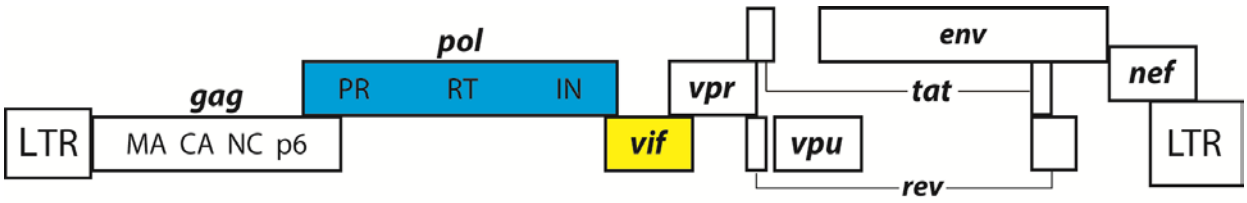


Figure 1. HIV-1 genome.

1.3. HIV replication

All the viral proteins are encoded as a single proviral transcription unit under the control of the HIV enhancer and promoter, called the long-terminal repeat (LTR). The LTR contains binding sites for numerous host factors that mediate the recruitment of the host RNA polymerase II machinery and cellular proteins that support low-level, poorly processive transcription. This basal transcription mainly gives rise to short (~60 base pairs (bp)) and thus incomplete viral transcripts, as well as a few full-length early transcripts [32]. Efficient, processive HIV transcription, which is crucial for viral replication, additionally necessitates the interaction of Tat with the transactivation response element (TAR) in the LTR and the host positive transcription elongation factor b (P-TEFb). Synthesis of Tat from rare full-length early transcripts fuels an auto-stimulatory transcription loop that sustains high levels of genome replication (reviewed in [33]).

A mature HIV virion consists of a lipid envelope, derived from the cytoplasmic membrane of the producer cell, and viral glycoproteins, enclosing a conical capsid that contains two copies of the viral RNA genome, as well as RT, IN, PR, Vif, Nef, and Vpr molecules. HIV enters a target cell via the engagement of the viral glycoproteins on the surface of the virion with its main cellular surface receptor CD4, and one of its chemokine co-receptors, CCR5 or CXCR4, which enables fusion of the virion's lipid envelope with the target cell's plasma membrane and release of the viral capsid into the cytoplasm. The co-receptor usage of a given HIV strain permits its classification as either R5- or X4-tropic, and more importantly, dictates the permissiveness of CD4+ lymphocyte subsets to the virus [34]. Other molecules have been implicated in viral entry, however their relevance *in vivo* is debated (reviewed in [35]). After entry, the viral genome is reverse transcribed into double-stranded DNA and imported along with IN and host factors within the PIC into the nucleus, where the proviral DNA is then preferentially integrated into actively transcribed host genes [36]. The fate of the integrated provirus is then largely dictated by the activation state and availability of host transcription factors in the host cell [37]. If the necessary host transcription initiation factors are available for proviral transcription, such as NF-κB, NFAT and AP1, replication of the provirus takes place, i.e. viral proteins and new viral RNA genomes are synthesized and assembled at the host cell membrane into virus-like particles, that become infectious (mature) virions after budding and PR-mediated activity (reviewed in [38]). The budding of progeny virions may lead to cell-lysis. However, innate immune sensing of viral nucleic acids, recognition and lysis by cytotoxic T lymphocytes (CTLs) and the cytotoxic effect of viral proteins are the main mechanisms associated with death of infected cells [39, 40]. As a result, generally, productively infected cells have a short life span of about two to four days *in vivo* [41, 42]. This means that viral burden in an infected individual is dynamic and involves continuous rounds of *de novo* infection and replication, offset by a rapid turnover of productively infected cells and cell-free virus (i.e. virions), due to viral cytopathic effects and clearance by HIV-specific CTLs [41]. Alternatively, the provirus may become inactive after integration and thereby persist in a non-immunogenic, latent state in the host cell (discussed

in more detail in section 1.5). Virus production and latency can be seen as alternative and dynamic states of replication-competent proviral HIV DNA, dictated by HIV's cellular tropism and dependence on specific cellular host factors for replication. In their productive state, proviruses ensure sustained viremia as well as viral evolution, dissemination and transmission. Whereas in their latent state, proviruses allow for immune evasion and long-term survival in the host, serving as archives of viral quasispecies that may resurface when conditions are fitting.

Viral replication is commonly measured by quantification of viral RNA by qPCR, or the viral p24 antigen by ELISA, in a HIV-infected sample (e.g. blood from a HIV infected patient or cell culture supernatant) [43-45]. In the clinics, a patient's viral load is quantified using the former method, whereby viral RNA is isolated from patient plasma (i.e. cell-free plasma virus), reverse transcribed and then quantified by qPCR, providing a value for plasma viral load in copies per milliliter.

As mentioned earlier, HIV requires the human CD4 receptor and CXCR4/CCR5 co-receptor for entry into the cells, so its cellular tropisms is intrinsically limited to cell subsets that express these surface receptors. However, cells that allow HIV entry may not be permissive to viral replication, either due to blocks imposed by innate immune mechanisms or the cell's activation state [37, 46]. Indeed, studies analyzing HIV DNA in peripheral blood mononuclear cells (PBMCs) from viremic HIV-patients revealed that the majority of infected cells harbor unintegrated or defective viruses [47], with <0.01% harboring replication-competent HIV [48]. Yet the relevance of cells harboring replication-incompetent virus for HIV pathogenesis and persistence is disputable [49, 50]. Owing to their permissiveness to productive infection, CD4+ T lymphocytes and macrophages are the main targets and producers of HIV. Other cell types that have been shown to harbor HIV *in vivo*, include monocytes, microglia, dendritic cells, Langerhans cells and hematopoietic progenitor cells (reviewed in [35, 46]). The main sites of viral replication are the (secondary) lymphoid tissues, including gut-associated lymphoid tissue (GALT) [51], which are also compartments harboring high frequencies of CD4+ T lymphocytes and macrophages. As different tissues are characterized by different cellular, immune and structural environments, and ARV drug distribution and metabolism are tissue-specific, viral diversity and replication can vary substantially between tissues and cells types, which is referred to as HIV compartmentalization [52, 53].

CD4+ T lymphocytes consists of a heterogeneous population of cells, with distinct transcriptional profiles, CXCR4/CCR5 expression levels and homing capabilities, which are dependent on their antigen-experience, activation and metabolic states, as well as anatomical compartmentalization [34, 54, 55]. In addition, HIV DNA and RNA are differentially and dynamically distributed across CD4+ T cell subsets and anatomical compartments [56, 57]. Namely, activated effector CD4+ T cells are the most amenable to reverse transcription and proviral gene expression and are the main producers of HIV in acute infection, whereas in chronic infection viral entry and replication are

predominantly observed in activated central or effector memory subsets [48, 57-59]. In addition, specific memory subsets have been identified as major sites of viral replication in particular tissues, such as T follicular helper (Tfh) cells and T helper 17 (Th17) cells in lymph nodes and gastrointestinal tract, respectively [60, 61]. On the other hand, infection of and replication in resting (both memory and naïve) CD4+ T cells is rare and inefficient [57, 62], owing primarily to blocks to viral entry and reverse transcription, as well as the presence of the host restriction factor SAMHD1 (reviewed in [63]). Furthermore, the ability of macrophages to sustain long-term productive infection and migrate throughout the body, as well as their longevity, make macrophages key players in HIV pathogenesis and likely physiologically relevant components of HIV persistence [64]. In non-lymphoid compartments, such as the brain, viral load is predominantly sustained by macrophages [65].

1.4. Combined antiretroviral therapy

ARVs have turned HIV-infection from a death-sentence into a manageable chronic disease, as they can effectively block viral replication and *de novo* infections, thereby curbing disease progression. Different classes of ARVs have been developed over the years, targeting key steps of the HIV life cycle. The main ARVs used in cART are: i) Nucleoside reverse transcriptase inhibitors (NRTIs) and non-nucleoside reverse transcriptase inhibitors (NNRTI), which interfere with reverse transcription by inducing chain-termination when incorporated by the RT, or by direct allosteric binding to the RT, respectively. ii) Integrase strand transfer inhibitors (INSTI), which block proviral DNA integration by rendering the IN catalytically inactive. iii) Protease inhibitors (PI), which hinder virion maturation by blocking proteolytic cleavage, resulting in the production of non-infectious virions. In addition, virion fusion inhibitors and co-receptor agonists (i.e. entry inhibitors) are other classes of less common ARVs. cART is made up of at least two classes of ARVs, such that multiple steps of the viral life cycle are simultaneously targeted, reducing the risk of emergence of drug-resistant strains [18].

The decay of viremia to undetectable levels (typically 20-50 copies/mL with ultrasensitive assays [45]) resulting from cART observed in treatment-naïve HIV-patients occurs in two phases [41], reflecting the contribution of different infected cell populations to viral production. The first phase is characterized by a rapid (exponential) decline in viremia within two to four weeks of treatment initiation, attributable to the clearance of cell-free plasma virus and virus-producing CD4+ T cells, which have half-lives of about seven hours and two days, respectively [41, 42]. The second phase is distinguished by a slower decay in viremia resulting from the elimination of productively infected cells with longer half-lives, such as macrophages, or with a lower activation state, such as partially activated CD4+ T lymphocytes [56].

1.5. HIV latency

1.5.1. Characteristics and establishment of HIV latency

The crux of HIV is its ability to exist in a very small portion of cells in a reversibly non-productive state, known as HIV latency. cART brought great hope for HIV eradication, as the rapid virologic control resulting from therapy suggested viral pools could be eliminated within 3 years of effective therapy [66]. However, the recovery of replication-competent HIV in long-lived resting memory CD4+ T lymphocytes of aviremic patients [9-11], accompanied by the detection of archival and non-evolving low-level residual viremia [67], highlighted the persistence of HIV despite prolonged viral suppression. Ultimately, the rebound in viremia in patients within weeks of cART interruption explicitly proved the presence of a latent HIV reservoir [13, 16], and revealed that cART cannot cure HIV. HIV persistence is achieved by the seeding of a latent reservoir in diverse cell types and anatomical compartments early after transmission [12, 13], that can be induced to produce infectious particles upon antigen recognition or exposure to T-cell-activating stimuli, and can thereby re-establish a disseminated infection in the absence of cART [10, 68]. Latently infected cells produce only low levels of viral RNAs and no viral proteins, making these cells immunologically indistinguishable from uninfected counterparts and therefore impervious to host immune surveillance [69-71]. Despite reports of markers that identify CD4+ T cell subsets enriched in HIV DNA in aviremic patients, such as CD2 [72] and CD32a [73], there is still no ubiquitous phenotypic method to unequivocally identify latently infected cells. Furthermore, studies in non-human primates (NHP) and humanized mice or with patient samples revealed that the HIV reservoir is essentially unperturbed by existing cART and cART intensification, LRAs and broadly neutralizing antibodies (bnAbs) [11, 74-79]. The main reservoir is found in resting memory CD4+ T lymphocytes in lymphoid tissues and has been evaluated to occur at a frequency on the order of one in a million resting CD4+ T lymphocyte [9, 11, 48, 51]. For this reason, most research on HIV latency and latency models are focused on this cell type. Macrophages have also been identified as reservoirs of HIV persistence, along with other cell types such as dendritic cells, monocytes and microglia, yet whether they persist under cART and their relative contributions to viral rebound upon cART interruption remain uncertain and controversial [64, 65, 80-84]. Due to the intrinsic longevity of memory CD4+ T lymphocytes, and the stable integration of HIV DNA in its host, the HIV reservoir is incredibly stable, with an estimated half-life of about 44 months [74]. Thus, despite the latent reservoir being very small, its stability and imperviousness to cART and anti-viral immune mechanisms, make it the major barrier to HIV eradication.

Strikingly, long-lived latently infected lymphocytes can sustain the HIV reservoir through clonal expansion and homeostatic proliferation without viral production [85-87]. The homeostatic maintenance of latently infected cells has important implications for eradication strategies aiming at purging the reservoir and underlines the difficulties that lie ahead with developing such strategies,

especially with regards to the required depth of purging and expectations linked to decay of the reservoir. Furthermore, in depth characterization of residual viremia (i.e. trace levels of cell-free plasma virus) in cART treated patients revealed that a homogenous population of plasma clones predominated, suggesting that under cART, virus is released due to activation of a small number of latently infected cells [88]. These predominant plasma clones (PPC) were rarely found in resting CD4+ T cells from the blood [88], implying that the circulating pool of resting CD4+ T cells, or any blood cell type, is unlikely to be the source of these PPC. These findings emphasize how viral and lymphocyte dynamics, as well as cellular reservoirs in the blood compartment, may not be representative of those in lymphoid tissues [51, 82, 89-94], but also that mixing and spreading can occur between compartments; highlighting the importance of studying HIV latency beyond the blood compartment.

The stable integration of retroviral genomes into the chromosomes of the cells they infect is an obligate step of retroviral life cycle and intrinsically confers the ability to persist for the lifespan of the cellular host. However, why and how HIV latency is established *in vivo* is not completely understood and still debated, especially as evidence is largely based on *in vitro* models. Clearly, HIV latency allows HIV to evade elimination by cART, but cART is a modern intervention and yet in rhesus macaques, latent SIV reservoirs are established within 3 days post-inoculation, before SIV-specific immune responses can be mounted [12]. Thus, latency must serve another, evolutionarily conserved, purpose. In addition, although the latent proviral landscapes in aviremic chronic HIV-patients are dominated (>98%) by CTL escape variants [69], ultimately HIV's capacity to rapidly evolve is what enables it to avoid recognition by neutralizing antibodies and CTLs [95]. Taken together these findings imply that escape from the host immune system may not be the key role of latency, but rather that HIV latency may be a deterministic evolutionary strategy to enhance viral transmission across mucosa (i.e. most common exposure route) [96, 97]. Specifically, long-lived latently infected cells could help HIV survive in its host during initial mucosal infection, a phase characterized by high rates of viral extinction [98]. In fact, it has been estimated that the probability of HIV-transmission is below 0.5% [99].

The establishment of HIV latency in resting memory CD4+ lymphocytes has yet to be fully elucidated, however two main mechanisms are favored. Either, HIV may directly infect and integrate its DNA into resting CD4+ T lymphocytes, which by virtue of their quiescent state result in silencing of the provirus [37, 100, 101]. Resting lymphocytes are largely refractory to HIV infection because their low CCR5 expression [34, 57, 62] makes for poor viral entry, and their low cytoplasmic deoxynucleoside triphosphates (dNTP) levels greatly limit reverse transcription of HIV RNA [102]. Nevertheless, infection of both memory and naïve resting CD4+ T lymphocytes has been achieved *in vitro* [103, 104], particularly in those with a memory phenotype or isolated from lymphoid tissues, where the microenvironment seems to skew these cells into a state more

favorable to HIV entry and integration [91, 105]. Alternatively, proviral latency is established in resting CD4+ T lymphocytes when productively infected activated memory precursor cells survive viral cytopathic and host cytolytic effects long enough to revert to a long-lived resting memory state [9, 25, 100]. This mechanism is particularly plausible given that activated CD4+ T lymphocytes are supportive of all steps of the HIV life cycle and that transition between different cellular activation and functional states, characterized by different gene expression profiles, is part of normal T cell memory generation [54]. Recently Shan et al. explicitly demonstrated that HIV infection can become latent as a consequence of the unique transcriptional state of CD4+ T lymphocytes undergoing effector-to-memory transition that makes these cells permissive to infection, but not conductive to viral production [57]. In either case, what appears to drive HIV infection into a latent state is the activation status (and associated gene expression profile) of the infected cell. Additional mechanisms may then play a role in enforcing latency.

1.5.2. Molecular mechanisms of HIV latency establishment and maintenance

A number of incompletely understood and overlapping molecular mechanisms, involving numerous host and viral proteins, mediate the establishment and maintenance of HIV latency, of which two forms have been observed: pre-integration and post-integration latency. Most of these mechanisms were discovered using *in vitro* models of HIV latency. Pre-integration latency arises due to defects in reverse transcription, inhibition of nuclear import of the viral PIC or incomplete integration of the proviral DNA, resulting in the formation of labile short-lived unintegrated forms of virus in the cytoplasm of infected cells [106]. As this form of latency is responsible for short-term latent reservoirs, it is of minor clinical importance [50, 107]. On the other hand, post-integration latency is responsible for the stable persistence of HIV in its host and results from a multitude of viral and cellular molecular mechanisms (reviewed in [108]). Briefly, the main mechanisms of post-integration latency establishment and maintenance (i.e. silencing of the provirus) involve transcriptional repression or post-transcriptional blocks. As discussed earlier, proviral transcription requires and is regulated by a number of inducible host transcription initiation factors (e.g. NF-κB, NFAT), whilst binding of Tat and cellular cofactors for HIV transcription (e.g. P-TEFb) to the viral promoter are required to relieve transcriptional blockade and achieve processive transcription. Thus, the cytoplasmic sequestration of key transcription factors, their presence in inactive conformations or inhibitory complexes, as is seen in resting cells (i.e. cells with low metabolic activity), and absence of Tat, prohibit efficient transcription. This underlines how the activation status of the infected cell greatly affects the fate of the provirus. Moreover, the site of proviral integration and chromatin environment of the LTR may lead to proviral silencing. Although HIV integrates non-randomly in actively transcribed genes [36], proviral latency may be reinforced by transcriptional interference during expression of neighboring host genes. In addition, the consistent positioning of two nucleosomes (nuc-0, nuc-1) in the LTR mediated by histone deacetylases

(HDACs), epigenetically restrict transcriptional elongation from the LTR [109]. Removal of this epigenetic block by hyperacetylation and remodeling of nuc-1 mediated by histone acetyltransferases (HATs), has been induced by Tat, NF-κB and various stimuli, including cytokines and phorbol esters [109]. Finally, post-transcriptional blocks affecting nuclear export, splicing and translation of viral mRNA may further suppress viral replication. For example, binding of mature host miRNAs to viral mRNAs can lead to their degradation or translational inhibition [110, 111], whereas viral miRNAs have been observed to contribute to latency maintenance via miRNA-mediated gene-regulation [112].

1.5.3. Measuring the HIV latent reservoir

Quantifying the HIV latent reservoir has been instrumental to identifying cellular reservoir [9], appreciating its stability [14], as well as understanding the effects of early cART [113], cART intensification [79] and latency purging strategies (discussed below), on the reservoir. As definitions of the HIV reservoir vary greatly, particularly depending on which forms of persistent virus are considered [50]. Herein we consider that the HIV reservoir denotes any cell type stably harboring replication-competent HIV DNA (i.e. intact proviral genomes) that can be induced to produce infectious viral particles. The last part of this definition is the pivotal characteristic of reservoir cells that makes them clinically relevant, yet it can be difficult to verify *ex vivo* [114]. Nevertheless, cells harboring unintegrated, defective, or uninduced persistence forms of HIV have also been implicated in HIV pathogenesis, evolution and persistence [49, 50, 114, 115].

Commonly reservoir size in aviremic patients is determined using *ex vivo* assays, performed on the patients' PBMCs or purified CD4+ T cells that aim to either quantify the frequency of cells that i) harbor HIV DNA; or ii) can be induced to produce HIV *ex vivo* [48]. The latter assay is commonly referred to as the quantitative viral outgrowth assay (qVOA), and is considered the gold standard for measuring the size of the replication-competent inducible HIV reservoir. Briefly, purified resting CD4+ cells from the peripheral blood of cART-suppressed patients are co-cultured in limiting dilution with permissive cells to propagate (amplify) released virus, and activated with phytohaemagglutinin (PHA) and allogenic irradiated PBMCs [48]. The common readout of the qVOA is p24 concentration in the culture supernatants, detected by ELISA. Over the years, a number of different VOAs have been developed using different permissive cells for co-culture or T-cell-activating stimuli and regimes [114, 116, 117], as well as alternative and more sensitive readouts to detect viral replication in the cultures [118, 119] (reviewed in [120]). The reactivated viral clones can also be harvested from the cultures for further characterization [86]. qVOAs have also been helpful in assessing the therapeutic potential of drug candidates to reduce the frequency of latently infected cells. On the other hand, measurement of cell-associated HIV DNA is achieved by highly sensitive and generally more standardized and simple qPCR assays. Importantly, the

qPCR assay design will impact what types of cell-associated HIV DNA are measured, such as total [115], integrated [121], or near-full-length [85] viral genomes and will also measure defective or replication-incompetent viruses. Accordingly, viral outgrowth and PCR-based HIV DNA assays provide different estimates of the frequency of latently infected cells in a patient, with the latter generally yielding estimates that are at least two logs higher [122] (reviewed in [123]).

1.5.4. HIV latency reversal

As mentioned earlier, latency reversal is the critical characteristic of latent proviruses that makes latency relevant to HIV pathogenesis and persistence. If a latent provirus lacks the ability to revert to a productive state, whereby it produces infectious virus, then post-integration latency would be a dead-end for HIV, with little overt clinical importance. Reversal of latency in patients on suppressive cART has been proposed as a strategy towards HIV eradication and major efforts are being devoted to finding latency reversing agents (LRAs) that can specifically reactivate latent HIV DNA and thereby purge the latent reservoir. This strategy is often referred to as “shock and kill” and the idea is to administer LRAs to patients on cART to activate virus production in latently infected cells, which may trigger their clearance due to viral cytopathic effects and CTLs, whilst cART protects against de novo infections [17, 124].

Various compounds have been proposed as putative LRAs owing to their biological activities associated with antagonism of mechanisms of latency maintenance (described in section 1.5.2). The main classes of LRAs include compounds that either alleviate epigenetic silencing of the LTR, such as HDAC inhibitors, or promote T cell activation, proliferation and signaling, such as protein kinase C (PKC) agonists, cytokines (e.g. IL-2, IL-7, TNF-alpha) and Toll-like receptor (TLR) agonists (reviewed in [125]). These compounds induce HIV transcription in latently infected cells *in vitro* or *ex vivo.* A limited number of LRAs have been tested in clinical trials, including the HDAC inhibitors valproic acid, vorinostat [77, 126, 127], panobinostat [128] and romidepsin [76], as well as disulfiram (thought to activate NF-κB) [129, 130], TLR9 agonist [131], bryostatin (PKC agonist, activates NF-κB) [75], cyclophosphamide (alkylating agent that interferes with DNA replication) [132], IL-7 [133] and several types of IL-2 based treatments (IL-2 alone, IL-2 plus IFN-gamma, IL-2 plus OKT3 antibody (reviewed in [134])). Except for bryostatin, all these compounds, induced measurable increases in cell-associated HIV RNA in the test subjects, nevertheless none resulted in a significant clinical benefit in that they either failed to reduce the size of the reservoir, or rebound in viremia after cART interruption was neither inhibited nor delayed. Of note, *in vivo* assessments of valproic acid’s effect on reservoir size have yielded contradictory results. In a first clinical trial, valproic acid was administered for three months to four patients on intensified cART and a small but significant decrease in the frequency of infected resting CD4+ T cells in the peripheral blood of three of the four subjects was observed by the end of the treatment [135].

However, since then no other *in vivo* study of valproic acid found an effect of this HDAC inhibitor on reservoir size [136-139].

Ultimately, research in this area seems to be hindered by the lack of a convenient, physiologically representative HIV latency model that would allow more telling screens of LRA candidates before transition to clinical trials. Moreover, the notion that to be effective, latency reversal strategies will also require additional interventions to get efficient clearing of reactivated reservoir cells, such as boosting of virus-specific immune mechanisms [140, 141], immunomodulatory agents [142, 143] or compounds that induce apoptosis [125], is gaining attention and being investigated. Purging of the entire reservoir is very likely to require combinations of LRAs targeting different mechanisms of latency maintenance [144]. Such insights and combinatorial approaches may lead to the development of effective eradication strategies. Nevertheless, in order to develop new therapies targeting the latent reservoir, it is imperative to further characterize the cells that harbor replication-competent latent provirus, elucidate their contribution to HIV persistence, as well as clarify the mechanisms governing latency establishment, maintenance and reactivation *in vivo*. The extremely low frequency of latently infected cells, the fact that these do not produce viral proteins and that no surface markers exist for identifying latently infected cells, make HIV latency very difficult to study. The situation is exacerbated by the limited availability of patient samples, largely confining the *ex vivo* study of HIV latency to the peripheral blood compartment, shallow lymph nodes, and sometimes cerebrospinal fluid and small intestinal biopsies, which as discussed earlier, may dramatically limit our understanding of HIV latency in the human body.

1.6. *In vitro* models of HIV latency

As HIV is unable to replicate in nonhuman cells and studying latently infected resting CD4+ T lymphocytes *ex vivo* is very difficult (see above), HIV latency has largely been studied using *in vitro* models. The various *in vitro* models of HIV latency (reviewed in [107, 145]) can be subdivided into two groups: chronically infected immortalized T cell lines and latently infected primary cells. Briefly, the first *in vitro* models of latency consisted of chronically HIV-infected immortalized T cell lines that constitutively express minimal amounts of HIV mRNA and proteins. Viral gene expression can be induced in these cell lines with mitogens or cytokines, providing the first indications of the contribution of LTR activity and chromatin environment, and proviral integration site to latency. However, ultimately the proliferative and clonal nature of immortalized cell lines do not recapitulate the quiescent nature of, nor the proviral integration site heterogeneity observed in resting primary reservoir cells. In addition, establishment of HIV latency in these cell lines was found to be a consequence of mutations in the viral genome or integration site, further questioning the relevance of these models for elucidating the *in vivo* mechanisms of HIV latency. Therefore, various groups developed more physiological models of latently infected resting memory CD4+ T cells by using

fresh primary CD4+ T cells from healthy donors. As it remains unclear whether *in vivo* latency in resting memory CD4+ T lymphocytes arises from infection of activated lymphocytes during effector-to-memory transition or direct infection of resting lymphocytes, models exist for both approaches. In addition, models differ in the types of primary CD4+ T cells, culture conditions and HIV variants used, which has given rise to a variety of primary cell models, representing different T-cell subsets, with different proviral silencing characteristics and reactivation responses [145].

To recapitulate latency establishment due to reversion of an infected activated lymphocytes to a resting state, the common approach involves activation of primary CD4+ T lymphocytes via T-cell receptor (TCR) stimulation and cytokines, infection and then culture for several weeks in the absence of T-cell stimulation to allow the infected cells to transition back to a resting state [146-150]. These models can yield high numbers of cells harboring integrated latent provirus [151], however the short life span of lymphocytes in culture in the absence of stimulatory signals is problematic. Thus, during the transition to a resting state, the primary cells are often cultured in the presence of low doses of cytokines (e.g. IL-2, IL-7, IL-15 [147, 152]) sufficient to promote cell survival without inducing T-cell activation. Alternatively, to maintain the infected primary CD4+ T lymphocytes in culture long enough for reversion to a resting state in the absence of T-cell stimuli some groups transduce the primary cells with anti-apoptotic proteins (e.g. Bcl-2 [148, 150]), or co-culture with feeder cells (e.g. H80 cells [149]). Possible caveats of these models is that the ensuing latently infected cells may not be truly quiescent as they continue to express markers of activation [149, 153], display an altered morphology [147] and often do not encompass all different CD4+ T cell memory subsets shown to harbor latent virus *in vivo* [107]. Alternatively, resting primary CD4+ T lymphocytes, which can consist of either purified or mixtures of naïve, central memory, and effector memory subsets, are directly infected *ex vivo* using chemokines to render the cells susceptible to infection [104], or spinoculation [103, 154], however the latter is not very physiological and may result in a substantial amount of cells carrying unintegrated genomes.

In addition, the genetic composition of the latent proviruses in the various models differ greatly, ranging from wild-type to HIV-derived reporter viruses, replication-competent to pseudotyped viruses. Replication-competent viruses encoding all or the majority of viral genes may result in more representative states of latency, but may reduce yields of latently infected cells due to cytotoxic effects, and require the use of ARVs in reactivation assays, endorsing the use of less physiological replication-incompetent or defective viruses. Furthermore, reporter viruses are advantageous in that they allow a simple means to distinguish productive and latent infection and quantify latency reversal. Yet, defective, pseudotyped and reporter viruses may display altered behaviors, affecting latency establishment, maintenance and reversal.

These primary models have been pivotal to the study of HIV latency and related mechanisms, yet these *ex vivo*-generated latently infected cells and the state of latency therein may not be entirely representative of the *in vivo* situation. Namely, the latently infected cells in these models may not be truly quiescent, nor reflect the phenotypic heterogeneity of reservoir cells found in patients, whilst the long-term *in vitro* culture involved may affect the physiology of the primary cells and chromatin environment of the provirus [107]. Indeed, an in-depth comparison of proviral reactivation in a panel of established latency cell models in response to thirteen stimuli known to reactivate latent HIV, revealed that no single *in vitro* model of HIV latency could consistently and accurately recapitulate the *ex vivo* response characteristics of latently infected resting CD4+ T lymphocytes from aviremic HIV patients [145]. The non-physiologically high proportions of latent infections that arise after *in vitro* infections (see Table 1 in [151]), further highlight a divergence from the *in vivo* scenario. Finally, *in vitro* models of latency cannot capture the complex interplay between HIV and the host immune system, which has been shown to affect and shape the latent reservoir in patients [6, 69, 97]. Therefore, there is the need for a physiologically relevant *in vivo* model of HIV latency, which would allow the isolation and study of genuine latently infected cells from different tissue compartments, and be used to test the efficacy of latency-reversing strategies in purging reservoirs systemically. To this end, humanized mice (hu-mice) engrafted with a human immune system provide a unique opportunity to study HIV systematically and in the context of a functional immune response. Additionally, hu-mice are exceptionally customizable models, as the human system can be gene-engineered prior to engraftment.

1.7. Humanized mice in HIV research

Hu-mice are immunodeficient mouse models engrafted with functional human systems, allowing the study of diverse human pathologies and used as pre-clinical models for testing new pharmacological agent. To this end, various immunodeficient mouse models have been generated via the introduction of mutations associated with immunodeficiencies. The latest generation of immunodeficient mice, called NSG (NOD-*scid* IL2Rgamma^{null}) mice, carry a mutation in *Prkdc*, causing the severe combined immunodeficiency syndrome (SCID), and knockout of the IL-2 receptor common gamma-chain, abrogating reactivity to many essential cytokines. NSG mice lack mature lymphocytes, functional natural killer (NK) cells and other components of the innate immune system [155, 156]. They are also thereby highly amenable to the engraftment of human CD34+ hematopoietic stem and precursor cells (HSPCs), from which a functional lymphoid system of human origin (including T and B lymphocytes, NK cells, monocytes, macrophage, dendritic cells) is reconstituted in the mice [155-158]. Humanized NSG mice can be generated in different ways based on the source of the HSPCs (i.e. cord blood, fetal liver and thymus, or bone marrow), their route of administration (i.e. intrahepatic, intraperitoneal or intravenous), age of the recipient immunodeficient mice (i.e. newborn or adult), as well as whether pre-conditioning is used

(reviewed in [159]). In our group, we transplant irradiated, newborn NSG pups intrahepatically with cord blood-derived CD34+ hematopoietic stem and precursor cells (HSPCs).

Hu-mice are cost and time effective, flexible models to study HIV latency (reviewed in [160, 161]), as they are highly susceptible to disseminated HIV-infection [162, 163], mount HIV-specific immune responses [164-167], harbor various tissue and cellular reservoirs (reviewed in [82]), which persist under cART regimes commonly used in the clinics and are resistant to immune clearance [69, 80, 163, 168-172]. New models of hu-mice, such as the T-cell only mice (ToM) and myeloid-only mice (MoM) ([161, 171]), are powerful tools for dissecting the contribution of different cell types to viral persistence and rebound after cART interruption.

1.8. HIV evolutionary potential

A critical feature of HIV is the accumulation of genetically diverse quasispecies during the course of infection [173]. This means that HIV exists in its host as a pool of incredibly diversified variants, rather than a homogenous viral population, responsible in part for the adaptive immune system's inability to mount an effective and long-lasting response, and failures in vaccine development [24, 25, 173]. HIV's incredible evolutionary potential is attributed to the low fidelity of its reverse transcriptase (RT) (3×10^{-5} mutations/base pairs/cycle [20, 174, 175]), high viral replication rates ($10^{10} – 10^{11}$ virions/day [21, 176]) and recombination between co-packaged genomes [22, 23]. Viral diversification in HIV patients is associated with increased fitness [177] and the rapid accumulation of variants resistant to recognition by CTLs and neutralizing antibodies, and ARVs [40, 42, 178], but can also result in the presence of drug-resistance variants prior to drug exposure [179].

Another proposed driver of viral diversification is sublethal mutagenesis induced by host cytidine deaminases from the APOBEC3 (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3) family. APOBEC3s contribute to human innate antiviral immunity and some of the family members have been shown to be cellular restriction factors of HIV, restricting HIV replication to varying degrees [71, 180-183]. These cellular deaminases hypermutate the nascent HIV cDNA during reverse transcription, which results in G-to-A mutations in the viral genome [184]. However, the viral accessory protein Vif can efficiently counteract this restriction, by targeting the APOBEC3s for proteasomal degradation [185, 186]. Strikingly, HIV variants encoding *vif* alleles with only partial-APOBEC3 neutralizing activities have been identified and recovered from HIV infected patients [187-189] and associated with ART-failure [189, 190]. Nevertheless, the contribution of the APOBEC3-Vif axis to viral evolution is highly debated [176, 191, 192], with opponents maintaining that APOBEC3-mediated mutagenesis results in lethal mutagenesis of the provirus (i.e. introduces premature stop codons or other mutations that render the provirus defective) or is insignificant compared to RT-error [176, 193, 194].

AIMS OF THE THESIS

The seeding of diverse cellular and anatomical latent HIV reservoirs throughout the body that persist under cART precludes HIV eradication with existing therapies. Strategies targeting the latent reservoir may enable the purging of persistent forms of HIV, and thereby eventually lead to cure of patients. However, specifically studying genuine latently infected cells seeded throughout the body is currently not feasible as these cells are largely phenotypically indistinguishable from their uninfected counterparts and access to patient samples is very limited and generally confined to the peripheral blood. These hurdles also limit our ability to effectively assess the efficacy of latency reversing pharmacological interventions. Indeed, despite inducing viral transcription in *in vitro* models of latently HIV infected cells, to date, all LRAs tested in clinical trials have failed to produce a measurable clinical benefit, highlighting the need for more representative models to screen putative LRAs. Therefore, there is a need for novel *in vivo* models of HIV that would allow the study of primary latently infected cells from diverse anatomical compartments and screening of novel therapeutic compounds.

The first project in my thesis aims to address this technological need. Specifically, the first specific aim of my thesis (chapter 2) is to **develop a HIV reporter hu-mouse model that allows the identification, isolation and characterization of primary latently HIV infected cells seeded throughout the host.** Such a model must incorporate a means to specifically and unambiguously identify, and thereby recover, the extremely rare reservoir cells. Humanized mice provide a HIV infection model that can easily be manipulated. Thus, we hypothesized that introducing a HIV-inducible reporter transgene into the NSG hu-mouse model would allow us to identify and sort genuine latently infected cells from diverse tissues in the hu-mice, based on their expression of the reporter gene. To this end, we collaborated with Professor Monsef Benkirane and Dr Gaël Petitjean (Institute of Human Genetics CNRS, Montpellier, FR), who had developed and validated an *in vitro* HIV fluorescent reporter system.

Another characteristic of HIV that greatly impedes its eradication is the virus's ability and propensity to diversify, helping it evade recognition by neutralizing antibodies and CTLs, and making vaccination strategies thus far ineffective. Although viral diversification is largely attributed to HIV's high replication rates and error-prone RT, APOBEC3 deaminases may also play a significant role in HIV diversification due to their mutagenic activity on the viral genome. The second specific aim of my thesis (chapter 3) is **to assess the contribution of APOBEC3G to HIV diversification and adaptability to an external selective pressure**. We hypothesized that as partially-active Vif alleles can be found in the circulating viral pool in patients and are associated with cART failure, suboptimal APOBEC3G neutralization *in vivo* is not fatal for HIV as a population and benefits viral adaptation. For this study, we collaborated with Professor Viviana Simon (Icahn

School of Medicine at Mount Sinai, New York, USA) who have expertise in sequencing viral quasispecies.

2. CHAPTER 2

Development of a HIV reporter hu-mouse model for studying HIV latency

2.1. Introduction

The rapid seeding of a very small, yet heterogeneous, disseminated and long-lived latent HIV reservoir, which is unaffected by existing anti-HIV therapies (cART, LRAs, bnAbs [11, 74-78]) and host immune surveillance [69-71], precludes complete HIV eradication. Consequently, HIV infection remains an incurable chronic condition affecting more than 36 million people worldwide (WHO, 2016 statistics). Efforts directed towards developing a cure for HIV can generally be divided into two categories based on whether they aim to achieve a functional or a sterile cure. The former aims at achieving durable virologic control in the absence of cART, such as fully blocking viral production from latent proviruses to inhibit viral rebound upon cART interruption, also known as the "block and lock" strategy. The latter entails purging the entire replication-competent inducible latent reservoir using LRAs together with cART, a strategy commonly referred to as "shock and kill". However, despite an extensive and growing body of research on HIV latency and the mechanisms that support viral persistence, a major barrier to effectively studying and targeting the HIV reservoir, irrespective of the approach pursued, is the lack of specific markers to identify all latently infected cells. Ultimately, this thwarts our ability to characterize latently infected cells, the provirus they harbor, as well as to understand the contribution of the different cellular and tissue HIV reservoirs to viral persistence.

Moreover, as latently infected cells are extremely scarce ($1/10^6$ CD4+ T lymphocytes) and access to human tissues is very limited, studying latently infected cells from patients is very difficult and circuitous. Currently the gold standard for quantifying the HIV reservoir in patients is based on *ex vivo* analyses of resting CD4+ T cells from suppressed HIV-patients, which are then cultivated and stimulated in limiting dilutions with feeder cells (QVOA) or analyzed by qPCR for cell-associated HIV DNA [195]. Such assays are useful for assessing the effectiveness of eradication strategies, exploring other cellular subsets that serve as HIV reservoirs [57, 85, 196, 197], and have helped reveal the pivotal role of clonal expansion of infected cells in HIV persistence [57]. Nonetheless, *ex vivo* studies are limited to analysis of tissue samples that can readily be obtained from patients, i.e. peripheral blood and shallow lymph nodes, and sometimes cerebrospinal fluid and small intestinal biopsies, whereas HIV persistence has been documented in a wide variety of anatomical site, in association with distinct and diverse cell types [82]. For example, Tfh cells in lymph nodes [198, 199], CD133+ hematopoietic progenitor cells in the bone marrow [200], CD4+ T cells as well as CD13+ myeloid cells and macrophages in GALT [82, 201-203], perivascular macrophages and microglia in the central nervous system [82], have all been found to harbor HIV DNA in aviremic treated patients. In this regard, the Last Gift Study, conducted at the University of California, San Diego, may provide unprecedented insights into HIV persistence in hard-to-reach anatomical compartments. The study aims to characterize the HIV reservoirs throughout the human body by rapid autopsy of full body donations from altruistic, terminally ill, HIV-infected individuals.

Although the value of such a project is evident, its implementation is too complex to fully replace the need for other approaches to study HIV latency. Moreover, for the rational design of new strategies effectively targeting persistent virus, qualitative characterization of genuine HIV reservoir cells and the latent proviruses they harbor is imperative. Using patient samples, this would ultimately require single-cell sequencing of millions of cells, as there are still no phenotypic means to unequivocally identifying latently infected cells. Therefore, there is a pressing need for physiologically relevant models of HIV latency.

In vitro models of HIV latency, which consist of either chronically HIV-infected immortalized cell lines that produce no or only very low levels of viral proteins unless stimulated, or highly manipulated primary cells infected with reporter viruses (reviewed in [107, 145]) (see thesis introduction, section 1.6), have been key to our early understanding of the molecular mechanisms governing HIV latency. However, such *in vitro* models cannot capture the heterogeneity of proviral integration [204], nor the quiescent nature of long-lived memory CD4+ T cells harboring latent provirus. In addition, the method by which latently infected cells are generated *in vitro* influences the nature of latency and the reactivation profile in these cells, limiting the relevance of results ensuing from such models. By measuring HIV reactivation in five primary T cell models and four latently infected T cell lines (J-Lat cell models) in response to thirteen known latency-reversing stimuli, Spina et al. demonstrated that ultimately none of the tested models could consistently recapitulate the *ex vivo* response characteristics of resting CD4+ T cells isolated from aviremic HIV-infected patients [145]. Finally, *in vitro* models cannot capture the complex interplay between virus and the host immune system. On the other hand, non-human primate models infected with SIV do not suffer from such limitations and pathogenic SIV infections closely approximate HIV pathogenesis in humans [205, 206], making NHP models valuable tools in HIV research. However, NHP studies are expensive, laborious, and not permitted in all countries, and ultimately do not allow the direct study of HIV in the context of a human immune system. In this regard, humanized mice models provide a powerful tool to study HIV (reviewed in [160, 161]). Namely, as hu-mice have a functional lymphoid system of human origin [156, 158, 207] (see thesis introduction, section 1.7), they are highly susceptible to HIV infection and recapitulate key features of HIV pathogenesis and persistence seen in patients [163, 168-170]. Moreover, hu-mice provide a highly mutable platform for studying HIV infection and persistence, as gene-engineered hu-mice (i.e. mice with a gene-engineered human immune system) are easily generated by lentiviral transduction of the human HSPCs used to reconstitute the immunodeficient mice [205, 208].

Therefore, in this study we sought to address the unmet technological need for a convenient, yet physiologically relevant model to study HIV latency by developing a HIV reporter hu-mouse model in which infected cells are specifically marked *in vivo*, thereby allowing genuine latently infected cells to be isolated for *ex vivo* characterization (Figure 1). The model is based on introducing an

inducible HIV infection reporter transgene into the human hematopoietic stem and progenitor cells (HSPCs) used to humanize immunodeficient NSG mice [207], which gives rise to HIV susceptible mice with human leukocytes that express a fluorescent protein upon infection. The HIV reporter system, called HR4lox, was developed and tested *in vitro* by Gaël Petitjean (Institute of Human Genetics, CNRS, Montpellier FR) and employs a Cre-lox recombination system to achieve infection dependent expression of the reporter transgene, red fluorescent protein (RFP). The Cre-lox system is a site-specific DNA recombination system derived from P1 bacteriophage and widely used in gene engineering and for conditional gene rearrangements, for example, to study gene functions, control gene expression and generate transgenic animal models [209, 210]. It involves two components, a Cre recombinase (Cre) and its specific DNA substrates, loxP sites. Cre is an enzyme that specifically recognizes two loxP sites and mediates rearrangement of the DNA that is flanked by these two loxP sites (commonly referred to as floxed DNA). If the loxP sites are in the same orientation, then the floxed DNA fragment will be excised, however if they are in the opposite orientations, then the floxed DNA will be inverted. Recombination is efficient in eukaryotic cells and does not require any accessory factors or specific DNA conformation [210]. Over the years, alternate loxP sites have been engineered, allowing for more complex gene rearrangements [211]. The HR4lox reporter system we use in our hu-mouse model is made up of two parts: a HIV reporter transgene containing a floxed RFP gene (HR4lox transgene) and a recombinant HIV that encodes a CMV promoter (CMV) and Cre cassette (CMV-Cre) that is called HOCC (Figure 1B). Specifically, the HIV reporter transgene contains a CMV-RFP cassette, in which the RFP gene is in antisense compared to the promoter and flanked by two sets of loxP sites (floxed-RFP), followed by a PGK promoter (PGK) and GFP cassette (PGK-GFP) (Figure 1A). This means that cells harboring the HIV reporter transgene (transgenic cells) are constitutively GFP+, but do not express RFP. RFP can only be expressed after two Cre-mediated recombination events take place at the loxP sites flanking the RFP gene. Thereafter, the upstream CMV promoter constitutively drives the expression of RFP. By using a HIV that encodes Cre for infections, in our model, Cre-lox recombination becomes dependent on HIV infection (Figure 1C). Thus, in our hu-mouse model, any transgenic cell infected with HOCC is specifically marked by its expression of RFP. This allows us to sort infected cells based on RFP expression using fluorescence-activated cell sorting (FACS). By treating HOCC-infected reporter hu-mice with cART, HIV latency and persistence can be explored in various deep tissues. We hypothesized that using this hu-model we could gain unprecedented access to genuine latently HIV-infected cells from a wide variety of tissues, allowing us to characterize these cells and gain further insight into the mechanisms driving latency establishment, maintenance and reactivation. We also envisaged using the model to evaluate the efficacy and potency of therapies aiming to purge the reservoir.

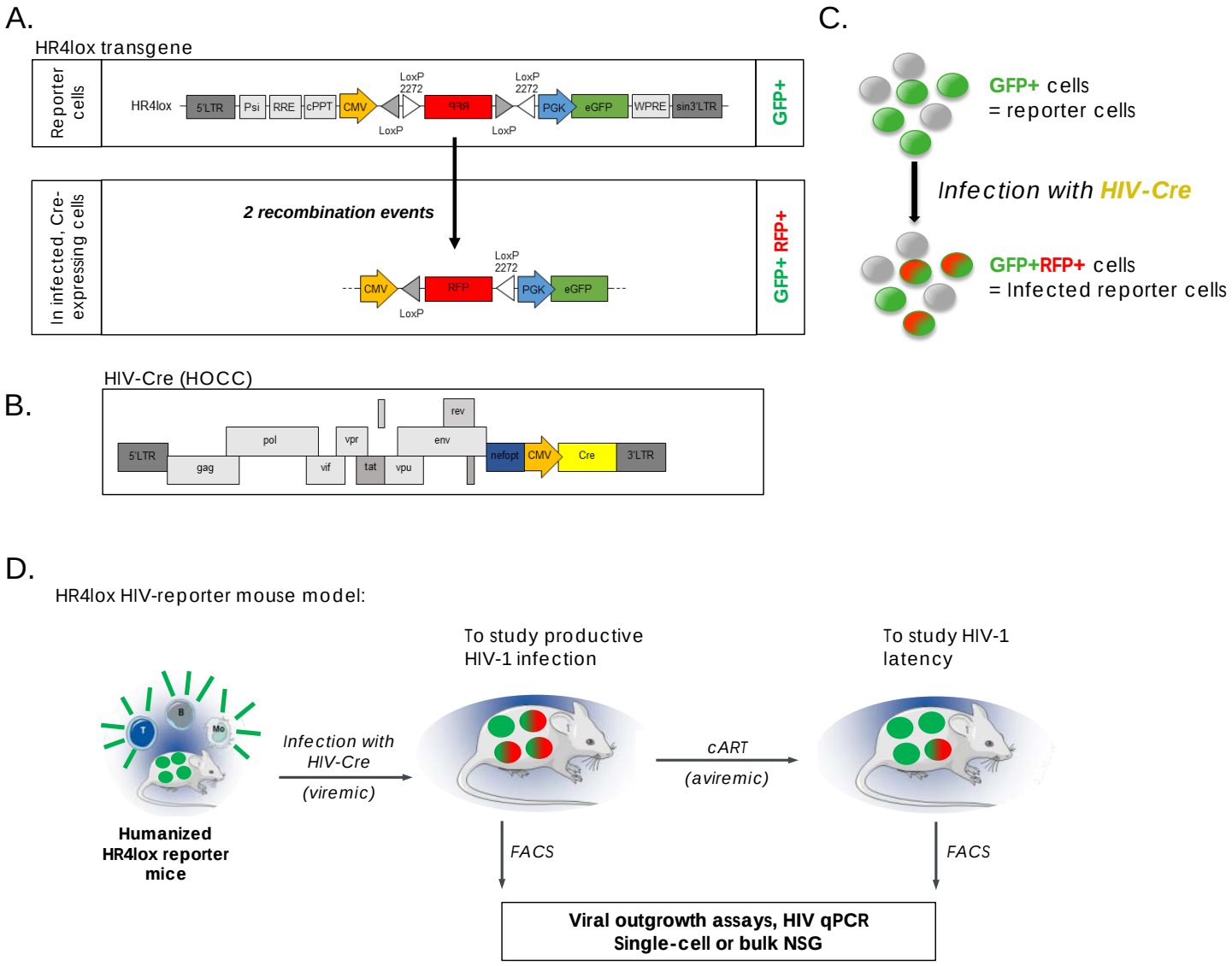


Figure 1. The HR4lox HIV-1 reporter mouse model and its components. A. *Schematic of the HIV reporter transgene, called HR4lox. Two Cre-mediated recombination events (inversion and excision) at two distinct sets of loxP sites are required for RFP expression.* B. *Recombinant HIV encoding a Cre, called HOCC, used for infection. HOCC is derived from the complete HIV (NL4-3 strain) sequence with a truncated Nef for optimized functionality and a CMV-Cre cassette at the 3' end.* C. *Schematic of the changes that occur when a cell carrying the HR4lox transgene is infected with a HIV encoding Cre. HR4lox transgenic cells constitutively express GFP from the PGK promoter and are only induced to express RFP upon infection, as only then is Cre present in the cells. Once Cre-lox recombination has occurred in a cell, RFP is constitutively expressed from the CMV promoter*. D. *Adult HR4lox hu-mice express the HR4lox transgene in CD34 lineage cells (represented as green cells) and upon infection with HOCC, these transgenic cells are induced to thereafter constitutively express RFP, allowing infected cells to be isolated by FACS based on the dual expression of GFP and RFP. To study the HIV reservoir, HOCC-infected HR4lox hu-mice are treated with cART and once they become aviremic, latently infected cells can also be identified and sorted based by gating on GFP+RFP+ cells. Parts A-B were adapted with permission from a figure made by Dr. Petitjean.*

2.2 Results

2.2.1 Generation of a HIV Reporter hu-mouse model

2.2.1.1. *In vitro* validation of the HIV reporter system

This work is based on a collaboration between the labs of Prof. Benkirane and Prof. Speck.
Dr. Petitjean assessed the functionality of the HR4lox reporter system *in vitro*, which we then implemented into hu-mice. For completeness, in this subsection I briefly summarize the results of these *in vitro* experiments.

As the HR4lox reporter system makes use of a genetically modified HIV strain encoding a CMV-Cre cassette, called HOCC, it was imperative to verify that this gene-engineered virus was genetically stable and remained functional over time during *in vitro* propagation [212]. Thus, firstly, the viability of HOCC and the stability of Cre expression were assessed by infecting MT4 cells expressing CCR5 (MT4C5 cells) and PHA/IL-2 activated PBMCs with the virus at a dose of 50-500 ng p24 per million cells and then longitudinally collecting cells from the cultures for flow cytometry and qPCR analyses. Detection of intracellular HIV p24 antigen by flow cytometry (Figure 2A-B) and quantification of cell-associated HIV gag DNA by qPCR both confirmed that the virus could effectively infect and replicate in either cell type, albeit at lower levels than WT HIV in the cell line for at least 4 weeks (data not shown). Expression of Cre in infected cells (i.e. p24+ cells) was also confirmed by flow cytometry analyses, but expression levels decreased during *in vitro* propagation. Quantitative PCR analyses revealed that *cre* was stably maintained in the HOCC genome during *in vitro* propagation in both cells types. Specifically, normalized to the proportion of infected MT4C5 cells or PBMCs, the average proportion of cells harboring *cre* was 80.86 ± 3.45% and 82.02 ± 6.72%, respectively, with no significant differences between cell types (p= 0.3962) (Figure 2C). These results confirmed that modification of the HIV genome did not abrogate HOCC's ability to infect and replicate *in vitro* and that Cre was expressed in infected cells. However they also highlighted that Cre's expression and stability are compromised over time, likely as a result of mutagenesis within *cre* arising due to selective pressures for removing all unnecessary gene content within the viral genomes during *in vitro* propagation driven by HIV's error-prone reverse transcriptase. These findings are in line with published work using recombinant HIVs [212].

Secondly, the functionality and specificity of Cre encoded by HOCC were verified by infecting MT4C5 cells that were transduced beforehand with the HR4lox reporter vector (HR4lox-MT4C5 cells), with HOCC or wild type (WT) HIV and then assessing GFP, RFP and intracellular p24 expression by flow cytometry (Figure 2D). Cre functionality was demonstrated by the induction of RFP expression in HOCC-infected cells. As RFP was only detected in HOCC-infected reporter

cells (i.e. GFP+) these results also confirmed the specificity of RFP induction in the reporter system. As expected, the majority of GFP+RFP+ cells stained positive for intracellular p24. The lack of p24 signal in a small portion of RFP+ cells could be a result of differences in the timing of viral gene expression compared to the Cre, whose expression is independently driven by the CMV promoter, (see Figure 1B), as well as some level of non-productive infection.

Lastly, it is crucial that Cre remains functional during serial passaging of HOCC to be of use for the *in vivo* model. Thus, Cre functional stability was assessed by co-culturing HR4lox-MT4C5 cells with MT4C5 cells or PBMCs that were infected with HOCC for 10 or 30 days, and then measuring RFP and p24 expression in the HR4lox-MT4C5 (i.e. GFP+) cells at 10 to 20 days post-infection (dpi). If the Cre in HOCC remains functional whilst replicating in MT4C5 cells and PBMCs, then infected HR4lox-MT4C5 cells should be induced to express RFP. RFP and p24 double positive HR4lox-MT4C5 cells could be detected in all co-cultures, indicating that the Cre encoded in the HOCC genome remained functional over up to 30 days of propagation *in vitro* (Figure 2E). These results are also depicted as the proportion of infected cells expressing RFP over the total proportion of infected cells in the co-cultures (i.e.: %RFP+ of p24+/ %p24+ cells) (Figure 2F), which provides a measure of the recombination efficiency in the reporter system. These results revealed that Cre maintained comparable levels of functionality when HOCC was propagated 10 or 30 days in either cell type ($p>0.05$). In other words, recombination efficiencies were stable and comparable over time ($p> 0.05$) when using HOCC passaged on MT4C5 (36.21 ±0.52%) or PBMCs (47.59 ±6.08%), yet efficiencies were significantly lower compared to that obtained with virus stock (65.59 ±4.92%) ($p<0.0001$), which is in line with the differences in Cre stability observed in the previous assays. Importantly, these co-culture assays highlighted that even *in vitro* the Cre-lox recombination efficiency of the reporter system is well below 100%. Nevertheless, taken together these results showed that the HR4lox reporter system was stable *in vitro* and could be used to specifically mark infected cells, prompting us to implement it *in vivo*.

2.2.1.2 Estimation of maximum readout obtainable with the HR4lox reporter system *in vivo*

To get an impression of the readout we can expect with the HR4lox reporter system *in vivo*, I used the Cre-lox recombination efficiency measured in PBMCs (i.e. 47.59%) to assess the proportion of infection events we can expect to detect via RFP, given different levels of reporter transgene expression in the HR4lox hu-mice (see Table 1). Namely, considering proportions of CD4+ T cells carrying the HR4lox transgene ranging from 10-100%, at best, about 4.7 to 47% of infection events occurring in the HR4lox hu-mice could be detected with the system, respectively (Table 1, column 2). Moreover, assuming 0.1-1% of CD4+ T cells are productively infected [6, 48, 213, 214] in the HIV reporter hu-mouse model, the number of CD4+ T cells that would need to be analyzed in order

to detect at least one infected cell was estimated. The results are summarized in Table 1 and reveal that depending on the level of transgene expression, between 2100 and 21000 CD4+ T lymphocytes need to be analyzed to detect a single infected cell based on the expression of RFP. Note that these are likely overestimations, as the true Cre-lox recombination efficiency in the reporter hu-mouse model is likely to be lower than observed *in vitro*. Nonetheless, this theoretical exercise highlighted the importance of achieving high levels of transgene expression in the hu-mouse model in order to maximize the readout, and also that if the recombination efficiency of the reporter system is too low *in vivo*, marking of infected cells may ultimately be too infrequent for effective isolation of latently infected cells.

		Given 1% infection rate	*Given 0.1% infection rate*	*Given 1/10^6 CD4+T cells are latently infected*
Proportion of CD4 T cells carrying reporter transgene (i.e. %GFP+ CD4+ T cells)	**% of infection events that can be detected**	**# of CD4+ T cells to analyse to detect 1 infection event**		**# of CD4+ T cells to analyse to detect 1 latently infected cells**
10	4.76	21'013	2'101	2.10E+07
15.5	7.38	13'557	1'356	1.36E+07
20	9.52	10'506	1'051	1.05E+07
30	14.28	7'004	700	7.00E+06
40	19.04	5'253	525	5.25E+06
50	23.80	4'203	420	4.20E+06
60	28.55	3'502	350	3.50E+06
70	33.31	3'002	300	3.00E+06
80	38.07	2'627	263	2.63E+06
90	42.83	2'335	233	2.33E+06
100	47.59	2'101	210	2.10E+06

Table 1. ***Estimation of reporter system readout in vivo.*** *All calculations assume an in vivo recombination efficiency of 47.59%, as measured in vitro in PBMCs.*

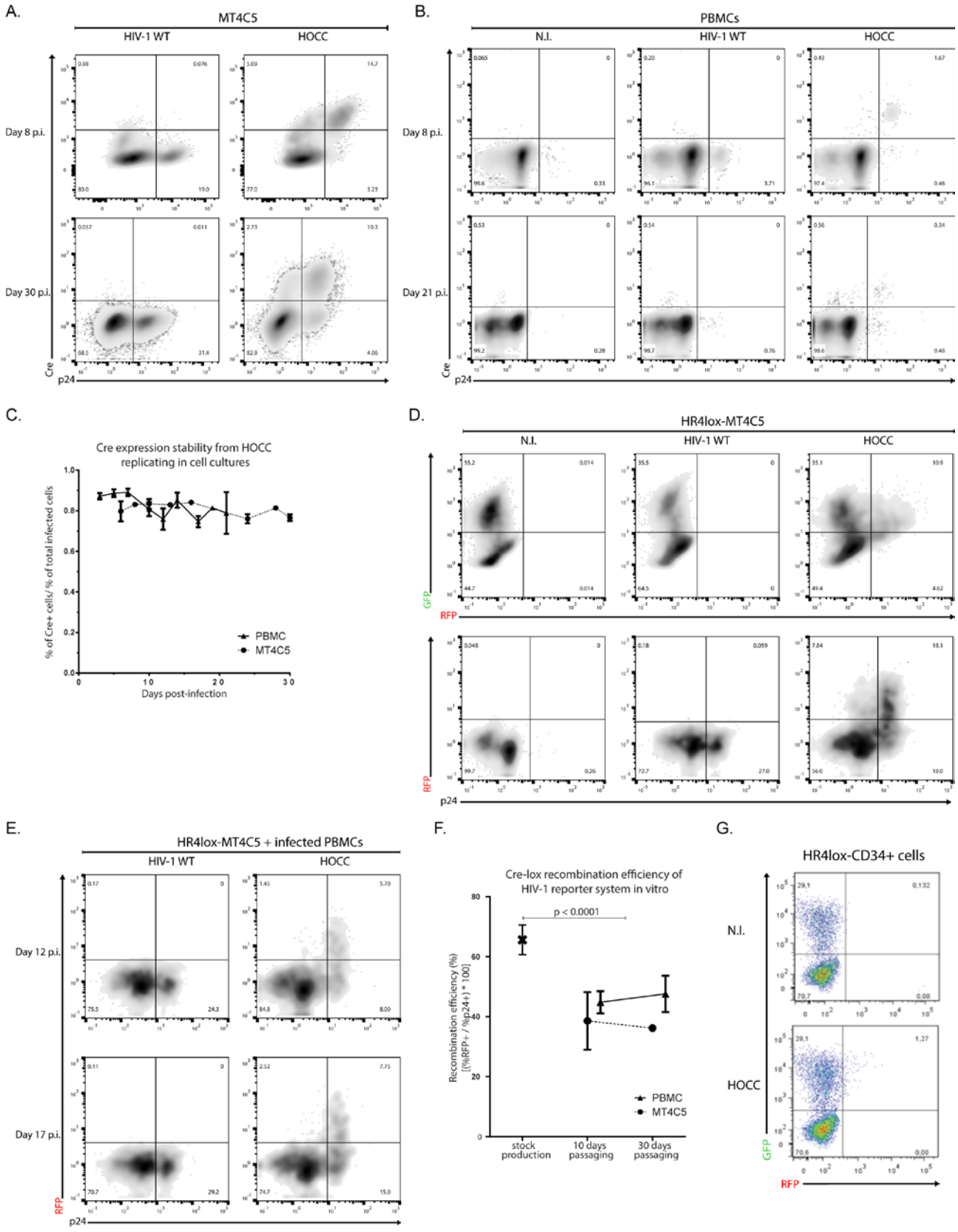


Figure 2. *In vitro* validation of the HR4lox reporter system and its components. A.-C. *Stability of Cre expression in HOCC-infected (p24+) cells: Flow cytometric analysis of (A) MT4C5 cells and (B) PBMCs infected with HOCC at 8 and 30 days post infection (dpi), showing that the majority of infected cells concurrently express Cre.* D. *Functionality of reporter system in MT4C5 cells. RFP expression is only observed in HR4lox-MT4C5 (GFP+) cells infected (p24+) with HOCC.* E. *Stability of Cre functionality in MT4C5 cells and PBMCs: Flow cytometric analysis of HR4lox-MT4C5 cells co-cultured with PBMCs infected with HOCC for 12 or 17 days.* F. *Recombination efficiency of reporter system in vitro, i.e. proportion of infected cells expressing RFP over the total proportion of infected cells in the co-cultures.* G. *Induction of RFP expression upon HOCC infection of CD34+ cells transduced with the HR4lox transgene. Figures based on data generated and collected by Dr. Petitjean.*

2.2.1.3 Implementing and testing the stability of the HIV reporter system *in vivo*

HR4lox hu-mice were generated as described in [163] using HSPCs that were transduced with the HR4lox transgene (see methods). Despite lentiviral transduction being well suited and widely used for the stable transgenesis of an extensive range of cell types, including non-dividing cells like HSPCs [215], it was imperative to verify that genetic engineering of the HSPCs did not compromise human immune system reconstitution, and if so, that transgene expression was stable over time. Flow cytometry analyses of PBMCs from three to four months old HR4lox hu-mice demonstrated that transduction of human HSPCs with the HR4lox lentivector did not impair human leukocyte reconstitution, as the engraftment of HR4lox hu-mice was comparable to that of regular hu-mice (i.e. transplanted with untransduced HSPCs) generated in our lab (Figure 3A). Moreover, as the HR4lox transgene constitutively expresses GFP, flow cytometry was used to assess the stability of the transgene *in vivo*. Flow cytometry analyses of cells from the peripheral blood, spleen and bone marrow of HR4lox hu-mice revealed that GFP was stably expressed from the HR4lox transgene in all probed anatomical compartments over time (Figure 3B), with GFP still detectable by flow cytometry in ten month-old HR4lox hu-mice. These results confirmed that the HR4lox-HSPCs were genetically stably and could efficiently support hematopoiesis in the immunodeficient NSG mice. However, the mean proportion of CD4+ T cells in the periphery carrying the reporter transgene was only 15.5±21.64% (range 0.13 to 83.80%); a consequence of low transduction rates (14.5 to 50%) of the HSPCs used to reconstitute the HR4lox hu-mice. Effectively, this low proportion of transgenic CD4+ T lymphocytes meant that on average, in the periphery, we would only be able to identify about one fifth of infection events. However, given that the *in vitro* recombination efficiency of the reporter system was estimated at 47.59%, the proportion of infection events we could expect to track in these hu-mice fell below 8% (see Table 1). As our goal was to use this HIV-reporter hu-mouse model to identify and isolate extremely rare latently infected cells (frequency on the order of one in a million resting CD4+ T cells), we needed to increase the detection rate of infection events. To this end, we subjected the transduced HSPCs to FACS, in order to obtain almost pure populations of HR4lox+ HSPCs for generating the reporter hu-mice. As can be seen in Figure 3A-B, the HR4lox hu-mice reconstituted with transduced and FACS-sorted HSPCs displayed higher proportions of transgenic cells. However, although HR4lox hu-mice generated from FACS-sorted HSPCs displayed normal human immune system development and comparable levels of humanization (i.e. %CD54+ cells) (Figure 3A and C), the engraftment of the FACS-sorted HSPCs occurred in an all-or-nothing manner, i.e. all hu-mice transplanted with the same FACS-sorted HSPC sample either developed a human immune system or did not (gray bars in Figure 3C). The overall engraftment efficiency in hu-mice cohorts generated from transduced and FACS-sorted HSPCs was 30.8%, and Fischer's exact test revealed this was a significantly lower proportion of successfully reconstituted mice compared to what we observed for WT (82.2%, $p<0.0001$) and solely transduced HSPCs (61.1%, $p=0.011$). Together these results indicated that

the HR4lox hu-mouse generated with a pure population of transgenic HSPCs constituted a stable reporter organism to use for the HIV reporter hu-mouse model.

2.2.1.4 Testing the stability and functionality of the recombinant HIV encoding CMV-Cre *in vivo*

Despite encouraging results of HOCC replication and functionality *in vitro*, the complexity of the human immune response to HIV made it imperative to verify that the recombinant virus retained its ability to replicate *in vivo,* whilst maintaining the Cre sequence in its genome. Infection of regular hu-mice with $4.4x10^5$ TCID50 HOCC revealed that albeit slower kinetics, HOCC replicated efficiently *in vivo* and was able to establish disseminated infections in seven out of eight hu-mice. However, persistent high level viremia (i.e. >1000 copies/mL) was only detected in five of these mice and only three mice had viremias that rose above 10^4 copies/mL (Figure 3D). Thus, in order to increase HOCC infection rates and resulting viremia at early time points [216], in subsequent experiments we used infectious doses of 1.0 to $1.2x10^6$ TCID50 per mouse. Furthermore, we confirmed by qPCR that the Cre sequence was maintained in the circulating viral pool, with Cre still detectable in plasma 8 weeks post-infection (Figure 3E). Together these experiments demonstrated that the recombinant HIV encoding Cre was suitable for use in the HIV reporter hu-mouse model.

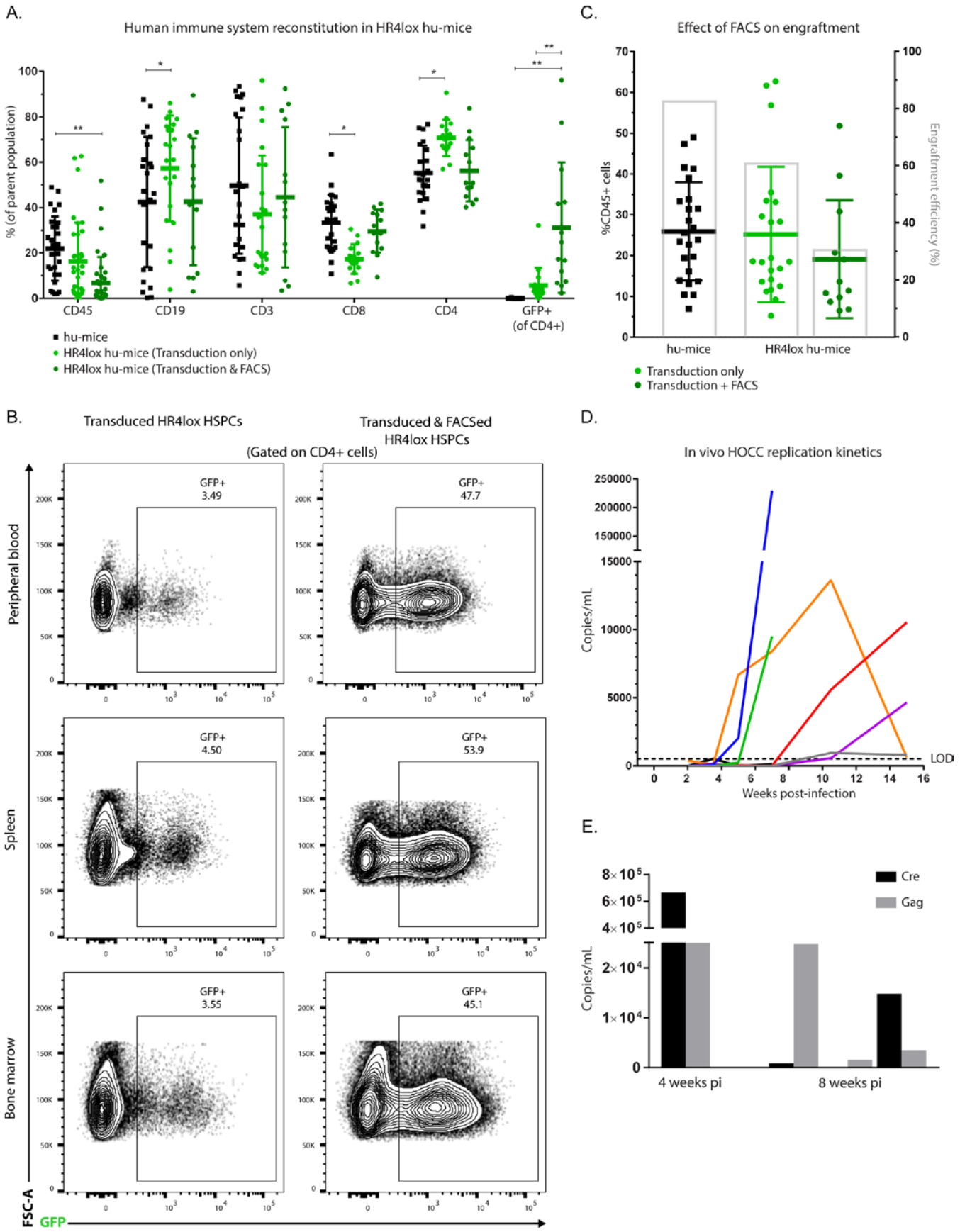


Figure 3. ***In vivo validation of reporter system components.*** A. *Flow cytometric analysis of human immune system reconstitution in HR4lox hu-mice compared to standard hu-mice. Transduction of CD34+ cells with the HR4lox transgene does not impair human immune system reconstitution as evidenced by comparable (2-way ANOVA) engraftment levels between the different groups of hu-mice. Note only hu-mice with ≥ 5% CD45+ are included in the subsequent downstream cell subsets. (Caption continued on next page.)*

Figure 3 (cont.). B. *Representative flow cytometry plots showing expression of GFP in CD4+ cells from peripheral blood, spleen and bone marrow of HR4lox hu-mice. GFP could readily be detect for up to 10 months.* C. *Human engraftment levels measured by flow cytometry in HR4lox hu-mice transplanted with HSPCs that had been only transduced or transduced and FACS-sorted, compared to standard hu-mice (i.e. HSPCs not transduced). The light gray bars correspond to the proportion of transplanted mice in which a human immune system reconstitution could be detected (i.e. ≥ 5% CD45+ in peripheral blood).* D. *Replication kinetics of HOCC in 8 standard hu-mice.* E. *Results from HIV gag and cre qPCR assays carried out on plasma viral RNA.*

2.2.2 Testing the functionality of the HIV reporter hu-mouse model: Proof-of-principle experiments in HOCC-infected HR4lox hu-mice

Having completed the *in vivo* implementation and validation of the HR4lox reporter system components, we proceeded to testing the functionality of the HIV reporter hu-mouse model; namely, whether HIV-infected cells could be expressly marked and detected in the HIV reporter hu-mouse model. As the frequency of productively infected cells is much higher than latently infected cells, as a proof of principle, we used a productive HIV infection scenario to test the model and we focused on CD 4 T lymphocytes, as these are the main cellular targets of HIV. To this end, cohorts of HR4lox hu-mice were infected with HOCC and once disseminated infection was confirmed by RT-qPCR, the animals were sacrificed and spleen, bone marrow, peripheral blood and, whenever possible, mesenteric lymph nodes were collected for analysis by flow cytometry, with a primary focus on GFP and RFP expression in these different compartments. In subsequent experiments, in order to characterize the RFP+ populations, we expanded the marker panel for flow cytometry to include phenotypic and activation markers.

Figure 4A depicts the gating strategy used to identify infected CD4+ T lymphocytes in the different tissues studied. In line with good flow cytometry practices, in order to set the cutoffs for the GFP and RFP gates, regular or HR4lox hu-mice infected with WT HIV were used as controls. Regular hu-mice have no reporter cells, and thus were expected to have no GFP or RFP positive cells, whereas HR4lox hu-mice infected with WT virus have GFP+ cells, but these were expected to not express RFP due to absence of Cre required for induction of RFP expression. When analyzing flow cytometry data, for reliability I only considered results from gates that included at least one hundred cells, with the exception of the RFP+ gate. In addition, as we noticed that with the first HR4lox hu-mice cohorts we sometimes had too few cells in the tissues samples (especially PBMCs and lymph nodes) to have a robust detection of RFP+ cells, thereafter and whenever possible we pooled corresponding tissues from several HR4lox hu-mice for the flow cytometry analyses. Thus, in the majority of cases the results presented herein are from pooled samples.

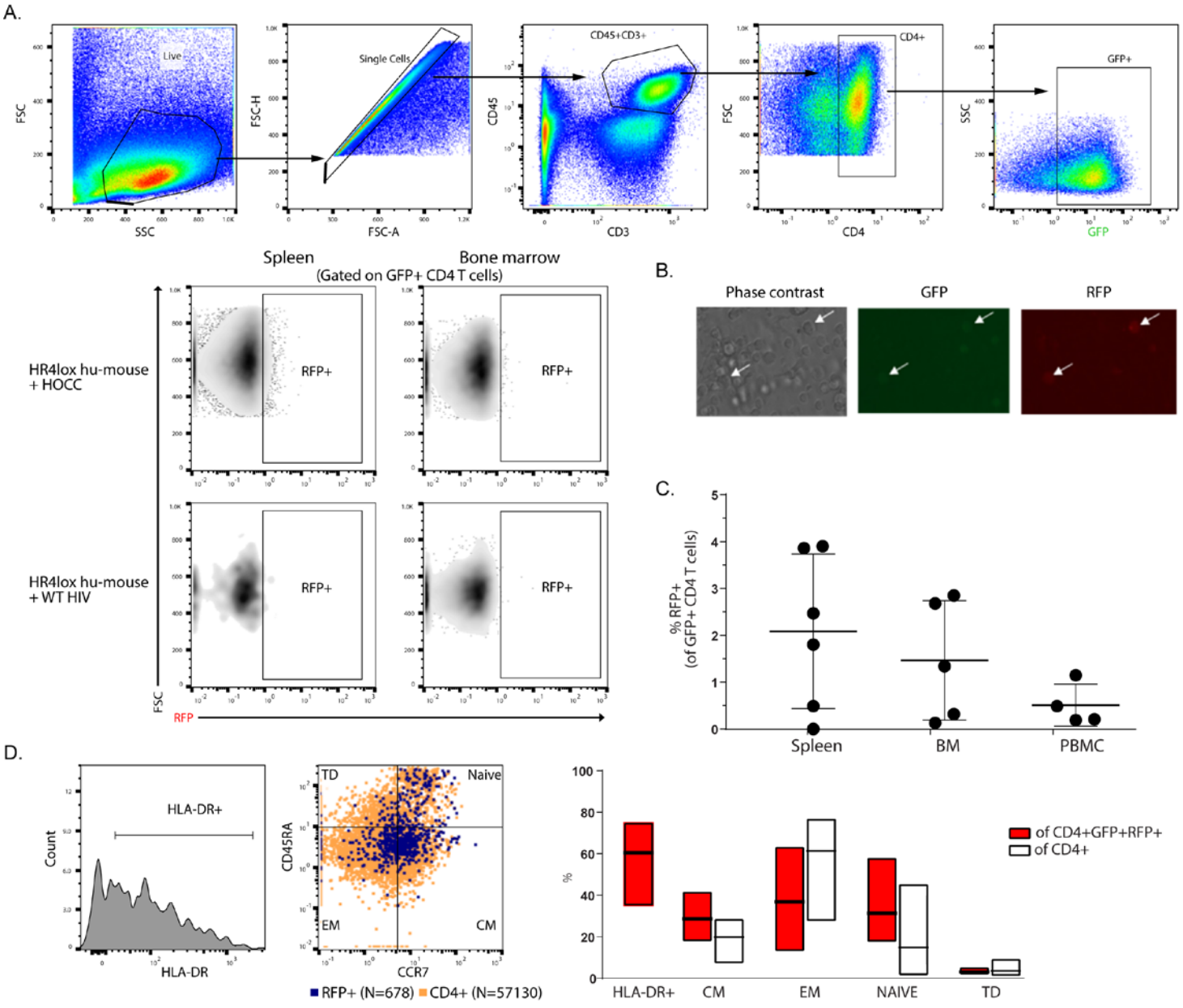


Figure 4. ***Productive infection in HIV-1 reporter mouse model.*** A. *Representative flow cytometry plots of spleen and bone marrow samples from HR4lox hu-mice infected either with HOCC or WT HIV. The upper panel shows the gating strategy used.* B. *Fluorescence microscopy images of splenocytes from a HOCC-infected HR4lox hu-mouse.* C. *Level of RFP expression in GFP+ CD4 T cells from different anatomical compartments of HOCC-infected HR4lox hu-mice.* D. *Phenotypic characterization of RFP+ splenocytes: representative flow cytometry plots are shown on the left and summary of results for all HOCC HR4lox hu-mice are shown in the right.*

Flow cytometry analysis revealed that GFP+RFP+ cells could be detected in the majority of HR4lox hu-mice infected with the Cre-virus, in all probed anatomical compartments (i.e. spleen, bone marrow, peripheral blood, lymph nodes), expect lymph nodes for which we only had very few samples of limited size. Representative flow cytometry plots for stained splenocytes and bone marrow cells of HR4lox hu-mice infected either with HOCC or WT HIV (NL4-3) are depicted in Figure 4A. In addition, fluorescence microscopy of a small aliquot of splenocytes from a HOCC-infected HR4lox hu-mouse confirmed expression of RFP in some of the GFP+ cells (Figure 4B). Marker expression derived from the flow cytometry analyses of the spleens of infected HR4lox hu-mice are presented in Figure 4 C-D. As expected, the highest frequencies of GFP+RFP+ cells were observed in spleens, where on average 2.09% of GFP+ cells also expressed RFP (max. 3.9%). In the bone marrows and PBMCs the average frequencies of RFP+ lymphocytes were 1.46% and 0.51%, respectively. These results translate to an average of approximately 20900, 14600 and 5100 infectious units per million (IUPM) GFP+ CD4+ T cells in the spleen, bone marrow and PBMC compartments respectively. Despite the relatively low fluorescence intensities of these RFP+ populations, RFP expression was only observed in conjunction with GFP positivity and RFP+ cells were not detected in uninfected HR4lox hu-mice or ones infected with WT HIV (see lower panels in Figure 4A). Furthermore, phenotypic characterization of total and GFP+RFP+ CD4+ T cells from infected HR4lox hu-mice samples by analysis of cell surface marker expression revealed that GFP+RFP+ CD4+ T cells from spleen, bone marrow, peripheral blood and lymph nodes were mainly activated cells (i.e. HLA-DR+) with a memory phenotype (i.e. CD45RA- [217]) (Figure 4D). These results are in accord with findings that in HIV infected viremic patients, or *in vivo* HIV/SIV models, productive infection mainly occurs in activated memory CD4+ T lymphocytes [60, 90, 197, 218-221], with relatively few productively infected cells at any given time resulting from their rapid clearance by host immune responses or viral cytopathic effects [21, 41, 89, 213, 214].

Despite low RFP intensities, together these results suggested that we were detecting infected cells marked *in vivo* rather than artefacts. Thus, we proceeded to isolating GFP+RFP+ cells from a new cohort of HOCC-infected HR4lox hu-mice by FACS, with the aim of using qPCR and a viral outgrowth assay (VOA) on these cells to confirm that indeed they harbored virus. As we expected a low yield of GFP+RFP+ cells from the FACS, in order to be able to run both assays in parallel with sufficient input material, whole genome amplification (WGA) was carried out on a small fraction of isolated cells destined for molecular analysis. The rest of the isolated cells were used for the VOA. The qPCR assay performed on amplified cell-associated DNA informs on the presence of the HIV *gag*, thus both defective and infectious proviruses are detected. On the other hand, the VOA informs on the presence of replication-competent virus via the detection of the viral p24 protein (a proxy for viral replication) in culture supernatant quantified by an ultra-sensitive digital ELISA (Simoa technology from Quanterix) [119], however its execution is more complicated and

laborious. These two complementary assays were implemented in order to robustly evaluate whether RFP+ cells in the model were truly infected cells (i.e. carrying integrated provirus). It is important to note that as WGA is not uniform [222], the subsequent qPCR is not quantitative and the CT values were used simply to assess whether a sample was positive for HIV-Gag.

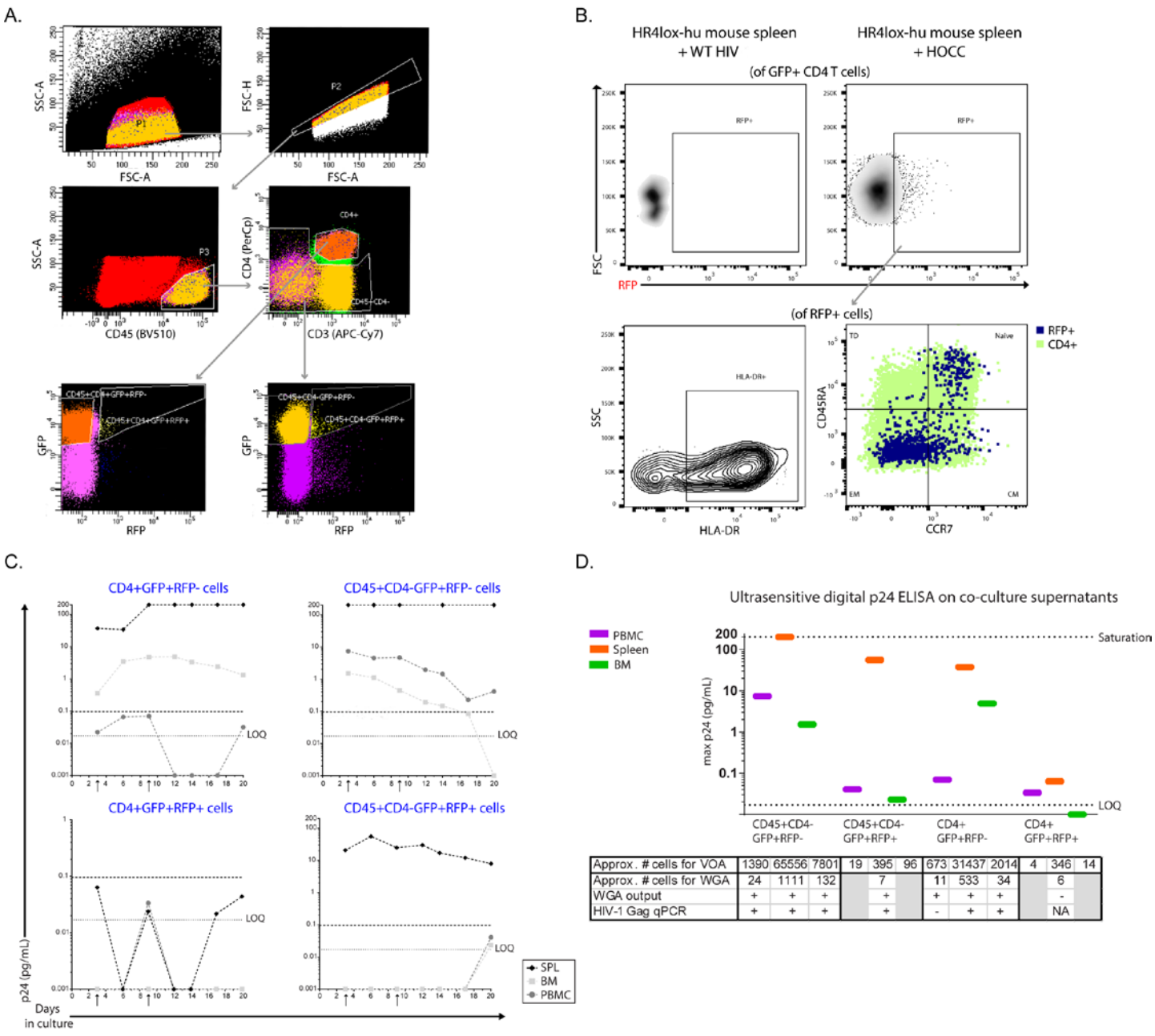
FACS of pooled spleen, bone marrow and PBMC samples from productively infected HR4lox reporter hu-mice allowed us to identify and phenotypically characterize GFP+RFP+ cells, as well as to collect these cells for further downstream analyses. Along with the main CD4+GFP+RFP+ target population, CD45+CD4-GFP+RFP+ cells were also isolated in order to capture any infected monocyte-lineage cells that could harbor virus in the reporter hu-mouse model. In addition, the GFP+RFP- counterparts for both of these cell subsets were collected to get an idea of the proportion of infection events we were missing due to ineffective recombination catalyzed by Cre or poor RFP expression. The sorting strategy and RFP expression plots are displayed in Figure 5A-B. Despite pooling of samples from several hu-mice and >66% of CD4+ T cells expressing the reporter transgene, as human engraftment levels were low, only 352, 14 and 4 CD4+GFP+RFP+ cells were isolated from the spleen, bone marrow and PBMC samples, respectively. The frequencies of RFP+ cells in the GFP+CD4+ T cell subset in these samples were 2.47%, 2.68% and 1.15%, respectively, and in line with our previous results and the literature.

The VOA on the isolated cell populations revealed that p24 production (i.e. viral replication), quantified with an ultra-sensitive single molecule array (SIMOA) digital ELISA (Quanterix Corporation) could be detected in all GFP+RFP+ sorted populations, with exception of CD4+GFP+RFP+ cells from bone marrow (Figure 5C-D). The ultra-sensitive digital p24 ELISA used for quantification of p24 production in the VOA has a limit of detection (LOD) of 0.0025 pg/mL [119] and thereby is >200-fold more sensitive than conventional p24 ELISAs [223], and has been employed to detect p24 production from a single productively infected CD4+ T lymphocyte [224]. Specifically, Passaes C. P. et al were able to demonstrate that p24 production from single infected cells range from 0.096 to 0.36 pg/ml, levels not quantifiable with conventional p24 ELISA [224]. For our analyses, as done in the study by Passaes C. P. et al, we chose a conservative limit of quantification (LOQ) of 0.017 pg/mL, which corresponds to the lowest concentration used for generation of the standard curve. Although the maximum levels of p24 produced by CD4+GFP+RFP+ cells isolated from spleens and peripheral blood were very low (0.033 to 0.063 pg/mL), the fact that few cells were used in the VOA and also that, at least for the splenocytes, the low p24 signal persisted over time – with the dips in p24 arising after each dilution of the well contents with fresh cells and medium – supported that this was not simply background noise. In addition, when large amounts of cells (n > 30000) were put in culture for the VOA, the p24 levels recorded in the culture supernatants were high (and sometimes above the range of quantification of the digital P24 ELISA), whereas with low input cell numbers, peak p24 levels in the

corresponding cultures were generally low. Furthermore, for the samples where DNA amplification was successful, the VOA results were corroborated by the Gag-qPCR results (see table in Figure 5D). Of note, despite having optimized the WGA protocol for a low number of input cells, in some cases no DNA was detected after the amplification reaction. As I did not have a method to verify that the desired fraction of isolated cells had effectively been transferred to the WGA reaction tubes, the assumption is that when no DNA could be quantified after WGA, this was due to the absence of cells in the input sample.

Viral replication and provirus were also measured in all the GFP+RFP- fractions, except in CD4+GFP+RFP- PBMCs where cell-associated HIV *gag* was not detected but p24 production was measured. The observation of viral replication and provirus in the RFP- subsets is not aberrant considering that Cre-lox recombination in the model is not 100% efficient, which entails that some infection events do not lead to the induction of RFP in GFP+ cells. However, these results revealed that in future sorts it would be pertinent to include an internal negative control population, such as CD19+GFP+ cells.

Taken together, these experiments validated that with the HIV reporter hu-mouse model we could identify, phenotypically characterize and isolate productively infected cells based on the dual expression of GFP and RFP. Thus, we proceeded to testing whether the model would also allow us to identify and isolate *in vivo* latently infected cells using the reporter genes.



Approx. # cells for VOA	1390	65556	7801	19	395	96	673	31437	2014	4	346	14
Approx. # cells for WGA	24	1111	132		7		11	533	34		6	
WGA output	+	+	+		+		+	+	+		-	
HIV-1 Gag qPCR	+	+	+		+		-	+	+		NA	

Figure 5. ***Isolation of RFP+ cells by FACS and subsequent characterization by VOA and HIV-gag qPCR.*** A. *Actual gating strategy during FACS*. B. *RFP expression in splenocytes of HR4lox hu-mice infected with HOCC versus WT HIV (upper panels) and phenotypic characterization of RFP+ cells (lower panels). The upper panel shows the gating strategy used.* C. *Kinetics of p24 levels in VOA culture supernatants for the four cell populations collected from HOCC-infected HR4lox hu-mice. The thick horizontal dotted line corresponds to the minimum amount of p24 produced by one infected cell (Passaes C.P., J Virol., 2017). Arrows indicate when activated healthy donor CD4+ T cells were added to the cultures.* D. *Overview of maximum p24 reading in the VOA and HIV-gag qPCR results after WGA for each isolated cell population. WGA was considered successful ("+") when DNA could be quantified by picogreen after the reaction and a sample. NA= not available.*

2.2.3. Studying the HIV latent reservoir in our HIV reporter hu-mouse model

To study HIV latency with the reporter hu-mouse model, viremic HOCC-infected HR4lox hu-mice were treated with a triple-drug ART until viral replication was suppressed below the limit of detection limit for a few weeks, as then, any GFP+RFP+ cell is a persistently infected cell that potentially contributes to the latent reservoir [168]. We assessed whether we could identify and isolate RFP+ cells persisting in different tissues from ART-suppressed HOCC-infected reporter hu-mice by FACS and subsequently subjected these cells to molecular and viral outgrowth assays to verify they harbored provirus. These experiments aimed to validate the specificity of the RFP marker in identifying HIV reservoir cells. The cART used herein consisted of two reverse-transcriptase inhibitors, Abacavir (ABC, 100 mg/kg per day) and Lamivudine (3TC, 150 mg/kg per day) and an integrase inhibitor, Dolutegravir (DTG, 15 mg/kg per day), dissolved in a sucralose solution replacing the normal drinking water [172]. This treatment regime was well tolerated by the animals and suppressed viremia below the limit of detection within 3-5 weeks. Moreover, gaining from our experience with the previous experiments, for these latency studies we modified the marker panel such as to include CD19 and thereby have a proper internal negative control. The new gating strategy used is depicted in Figure 6A.

In the first two cohorts of cART-suppressed HOCC-infected HR4lox hu-mice, GFP+RFP+ splenocytes could be identified and phenotypically characterized by flow cytometry, mainly in the CD4+ cell subset, but also in CD3-CD19- cells (Figure 6B), which likely correspond to HIV reservoir macrophages [80, 171]. Splenic and bone marrow RFP+ CD4+ T cells arose at a mean frequency of 0.204% (range 0.081 to 0.36%) and 0.680% (range 0.21 to 1.49%), respectively. Interestingly, the majority (~70%) of RFP+ cells in the hu-mouse model displayed a naïve phenotype (CD45RA+CCR7+), and to a lesser extent (~21%) central memory phenotype (CD45RA-CCR7+), whereas overall, effector memory cells were the most prevalent CD4+ T cell subset in the hu-mice (Figure 6B). This unexpected naïve phenotype may arise because humanized NSG mice do not fully recapitulate the T lymphocyte subsets found in human adults (see Table 2), and thus the higher prevalence of naïve T cells skews infection rates for this subset. Nonetheless, it would be of interest to elucidate what makes these CD45RA+CCR7+ cells susceptible to persistent infection. Alternatively, it could well be that the lymphatic reservoir is different from that found in the peripheral blood, especially considering that viral replication mainly occurs in lymphoid organs and GALT. Finally, new CD4+ T lymphocyte subsets are being identified as reservoirs, such as T helper follicular cells [60, 225, 226] and regulatory T cells (Treg) [218, 227]. Therefore the phenotype of the GFP+RFP+ cells in aviremic HR4lox hu-mice needs to be further evaluated by including additional markers in the staining panel, such as CD25 and CD127 (Treg), PD-1 and CXCR5 (Tfh).

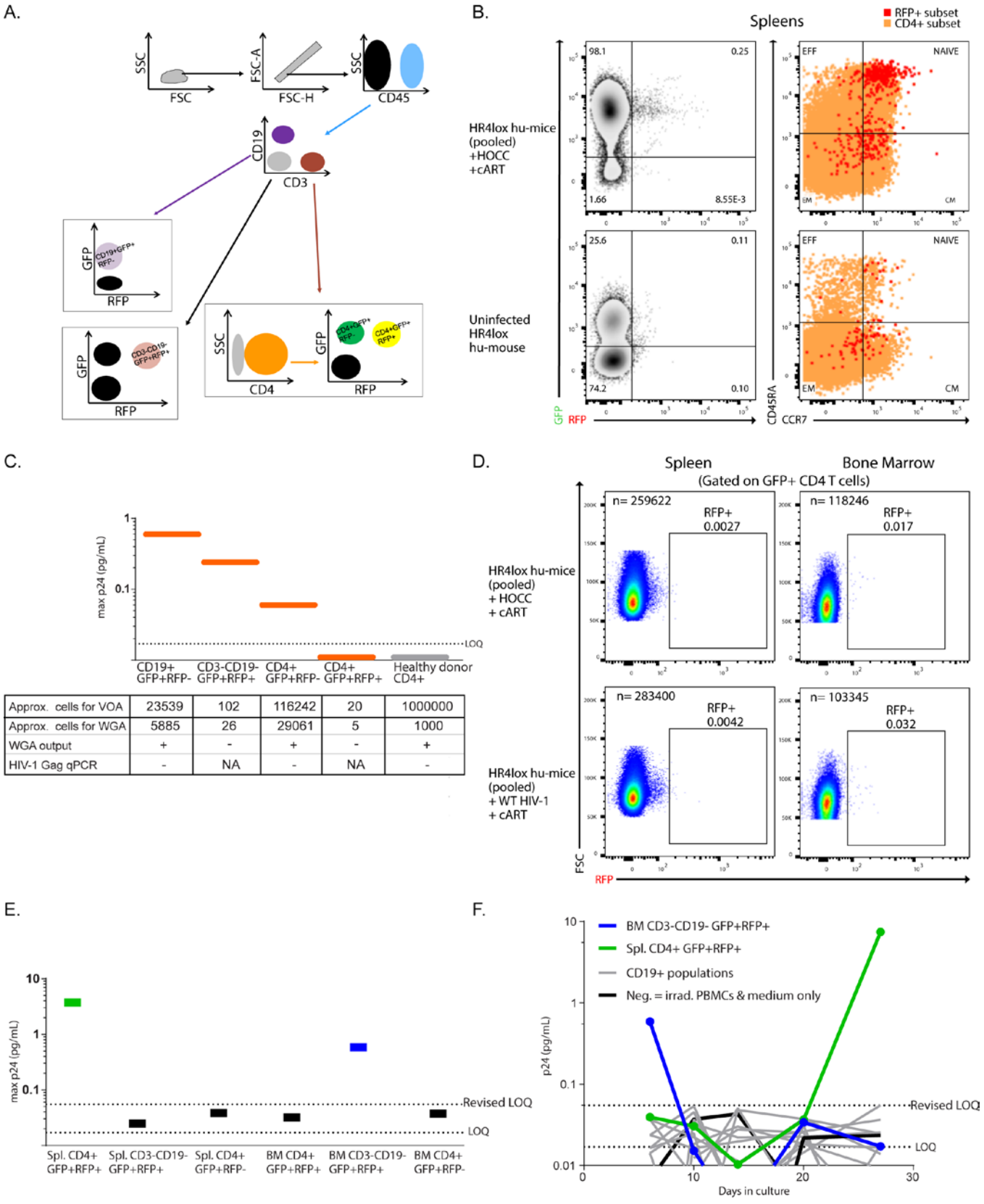


Figure 6. ***Exploring latency in the HIV reporter hu-mouse model.*** A. *FACS gating strategy for isolating latently infected (i.e. RFP+) cells*. B. *Representative flow cytometry plot of (left panels) RFP expression in splenocytes of cART-suppressed HOCC-infected HR4lox hu-mice compared to an uninfected counterpart and (right panels) phenotypic characterization of RFP+ cells. (Caption continued on next page.)*

Figure 6 (cont.). C. *Overview of* m*aximum p24 readings in VOA culture supernatants for the different cell populations isolated from cART-suppressed HOCC-infected HR4lox hu-mice (i.e. from sort shown in B).* D. *RFP expression in pooled samples from a new cohort of cART-suppressed HOCC-infected HR4lox hu-mice compared WT HIV-infected counterparts, when analyzing comparable numbers of cells.* E. *Overview of maximum p24 readings in the VOA culture supernatants for the different cell populations isolated from cART-suppressed HOCC-infected HR4lox hu-mice (i.e. from sort shown in D).* F. *Kinetics of p24 readings for the two VOA cultures that showed a single positive p24 read (blue and green lines), compared to the negative control (black) and the cultures containing the different isolated CD19+ cell populations (gray).*

VOA on the isolated cells revealed that inducible, replication-competent provirus persisted in the splenic CD4+ and CD3-CD19- cell compartments (Figure 6C), however not in the rare CD4+GFP+RFP+ fraction. When available, these results were corroborated by HIV *gag* PCR on a small fraction of cells subjected to WGA (see table in Figure 6C). Noteworthy are the substantial differences in the number of cells used as input to the VOA between the different isolated populations, with only about 20 CD4+GFP+RFP+ cells available for VOA, compared to more than 100'000 CD4+GFP+RFP- cells. Again, it must be noted that due to inefficient Cre-lox recombination, GFP+RFP- cells may harbor provirus. The lack of detectable viral replication in splenic CD4+GFP+RFP+ cells opened up a number of questions as to why we could not detect p24 in the culture supernatants. Namely, whether the reactivation stimuli used were insufficient to reverse latency in these cells [86, 114]; or whether reactivated latently infected cells only gave rise to abortive replication in the given culture conditions [224, 228]; or whether the isolated cells harbored defective provirus [195]. In addition, unexpectedly, the highest level of p24 production arising from VOA was seen for the CD19+GFP+RFP- splenocytes. The most plausible explanation for this was contamination during sorting arising as a result of the suboptimal anti-CD19 staining, which made distinguishing the CD19+ population from the CD19- difficult. Thus, it is possible that CD19- cells fell into the CD19+ sort gate. In addition, we realized that as we were pooling the HOCC-infected HR4lox hu-mice samples, we were actually analyzing a lot more cells for these samples than the control samples (i.e. tissue samples from a single hu-mouse), resulting in a greater amount of background noise in the experimental group compared to the control group. Therefore, we repeated the FACS of cells from a new cohort of cART-suppressed HR4lox hu-mice, a) using an optimized staining protocol yielding a better separation of the CD19 negative and positive populations, and b) acquiring an equivalent number of events for the control and experimental group samples. In addition, for the VOA, similar numbers of RFP+ and RFP- sorted cells were used for each population subset (Figure 6E). In this last proof-of-principle experiment aiming to assess if latently infected cells could be isolated from the HIV reporter hu-mouse model, flow cytometric analysis revealed that the RFP signal was too weak to detect persistently infected cells above background noise. This can be seen in Figure 6D from the lack of difference in RFP

signal between the control and experimental groups when equivalent number of events were analyzed. The frequency of GFP+CD4+ spleen and bone marrow cells that expressed RFP was only 0.003% and 0.017%, respectively. Furthermore, VOA on the isolated cell fractions did not give rise to any persistent p24 signal. We also observed p24 readings above the LOQ (0.017 pg/ml) for the negative controls; with the highest p24 reading from wells containing just irradiated PBMCs and medium at 0.043 pg/ml and the highest p24 reading arising from CD19+ populations at 0.05 5pg/ml (Figure 6E-F). Thus, using a revised threshold for p24 positivity, based on our negative controls, of 0.055 pg/ml, only CD4+GFP+RFP+ splenocytes and CD3-CD19-GFP+RFP+ bone marrow cells gave rise to a single positive p24 reading of 7.42 and 0.589 pg/ml, respectively. As in the case of the CD4+GFP+RFP+ splenocytes the positive p24 reading was measured at the very last assay time point, it is impossible to comment on whether this signal would of been sustained or have grown over time (Figure 6E). However for the CD3-CD19-GFP+RFP+ bone marrow cells, it is feasible that the signal arose due to p24 released from an insufficiently reactivated provirus or abortive viral reactivation, as the positive p24 reading appeared in the first sampling time point. Thus despite a low resolution, making the distinction of RFP+ cells from background noise very difficult, the analyses carried out on the isolated cells support that these cells contained replication-competent provirus. In any case, in hindsight the results from the first latency studies in the HIV reporter hu-mouse model depicted in Figure 6B-C are likely erroneous, and should not be considered.

	Frequencies in spleen (% of CD4+ T cells)			
	Human adult	**Human young adult (15-25yrs)**	**Human infant (2months to 2yrs)**	**Humanized NSG mouse**
Naive	17*	30	70	31*
Central memory	32*	20	10	32*
Effector memory	44*	45	15	19*
Effector	5*	2	2	3
Sources **average of values extracted from publications*	*Colineau L., PLoS One (2015); Sathaliyawala T., Immunity (2012); Langeveld M., Eur J Clin Invest (2006)*	*Thome J.J.C., Nat.* Med *(2016)*		*Audige A. BMC Immunol., (2017); Arainga M., Sci. Rep. (2016)*

Table 2. ***CD4 T lymphocyte subsets in humans compared to adult NSG hu-mice.*** *Data extracted from publications as indicated in the last column.*

2.3 Discussion and Outlook

2.3.1 Discussion

Although cART can effectively suppress HIV replication and thereby dramatically decrease morbidity and mortality of HIV-infected patients, it is not curative. The major barrier to cure is HIV's ability to hide in a latent form in various cell types and tissues despite years of successful cART. Upon treatment interruption, a disseminated infection is rapidly re-established from these latently infected cells. Consequently, cART must be taken lifelong, which is associated with numerous drawbacks, amongst which an elevated risk of emergence of drug resistances. In addition, a significant proportion of the HIV-infected individuals do not have access to cART [19]. Therefore, new therapeutic curative approaches are needed. Our knowledge of HIV reservoir cells remains fragmentary, primarily because of the difficulty in isolating latently infected cells from patients, which are only present at extremely low frequencies (~$1/10^6$ lymphocytes). Qualitative profiling of primary latently infected cells seeded throughout the human body may bring to light ways of specifically eliminating these cells, for example by effectively inducing the provirus they harbor (i.e. "shock and kill" approach), or ways to permanently block the latent proviruses (i.e. "block and lock" approach). As no ubiquitous surface marker exists that unambiguously identifies latently infected cells from a pool of healthy cells, such a characterization is currently not feasible. Indeed this would require that millions of cells from HIV-infected patients on cART be sequenced individually, and even then HIV reservoirs in hard-to-get tissues, such GALT or lymph nodes, would be missed. *In vitro* models of HIV latency are limited to the study of latency in specific cell types, fail to accurately recapitulate the quiescent nature of latently infected cells, rely on artificial methods of inducing latency, and cannot capture the complexities of host-virus interactions. Furthermore, NHP models of HIV latency are difficult to conduct, costly, time and labor intensive, and rely on the use of SIV or SHIV to model HIV. On the other hand, humanized mice offer a uniquely malleable model to study HIV in the context of a functional human immune system. Therefore, in this study we aimed to develop a HIV reporter humanized mouse model that would allow us to specifically identify and isolate genuine latently HIV infected primary cells from diverse anatomical compartments. Our model is based on a HIV infection reporter system (HR4lox reporter system, designed and tested *in vitro* by Dr. Petitjean), implemented into the well-established humanized NSG mouse model [207], such as to obtain reporter hu-mice in which infected cells can be identified based on the expression of the reporter gene, RFP (Figure 1). Explicitly, the HIV reporter transgene includes a dually-floxed and inverted RFP reporter gene downstream of a CMV promoter, which means RFP can only be expressed after two Cre-mediated recombination events have taken place (Figure 1A). This transgene is introduced by lentiviral transduction into the HSPCs used to humanize immunodeficient NSG mice. When a recombinant HIV that encodes a CMV-Cre cassette (HOCC, Figure 1B) is used to infect the transgenic hu-mice, infected cells carrying the reporter transgene (i.e. GFP+ cells) are induced to thereafter express RFP

constitutively, because the virus brings Cre into these cells. We hypothesized that with such a HIV reporter hu-mouse model we could map the sites of HIV persistence in different anatomical compartments and cellular subsets of the human immune system, as well as isolate genuine latently infected cells for subsequent genomic, transcriptomic and epigenomic profiling. To this end, we first validated the reporter system components *in vivo* and then tested the functionality of the HIV reporter hu-mouse model as a whole by infecting HR4lox-mice with HOCC and assessing whether a) RFP+ cells from tissue samples of viremic animals could be detected and isolated by flow cytometry and FACS; and b) the isolated RFP+ cell populations actually were infected using a viral outgrowth assay (VOA) and HIV-gag qPCR. Finally, we proceeded to explore HIV latency with our model.

Genetic manipulation of the HIV genome can result in viruses with impaired infectivity or dysfunctionality of the transgene [212]. Thus, although infection of HR4lox-MT4C5 cells with HOCC passaged on PBMCs for varying lengths of time demonstrated that the recombinant virus stably maintained its ability to induce expression of RFP in reporter cells *in vitro* (Figure 2E), the *in vivo* functional stability of the Cre recombinase encoded by the HOCC virus needed to be validated. Infection of standard hu-mice with HOCC resulted in detectable viremia in all but one animal, with sustained viremia in ~71% of the infected hu-mice (Figure 3D), confirming that the recombinant virus encoding Cre is infectious. In addition, Cre was detected in the circulating viral quasispecies up to eight weeks post-infection, implying the stability of the Cre within the HOCC genome during replication *in vivo*. We also verified that introduction of the HR4lox transgene into the HSPCs by lentiviral transduction did not compromise human immune system reconstitution in the resulting mice. Flow cytometric analysis of leukocyte reconstitution in HR4lox-mice, longitudinally in peripheral blood, and up to ten months post-transplantation in spleen and bone marrow, confirmed that the HR4lox transgene did not interfere with hematopoiesis (Figure 3A-B). However, as the frequencies of GFP+ human cells in our first cohorts of HR4lox-mice were low, which would reduce readout from the model, for subsequent humanizations, the transduced HSPCs were sorted by FACS to obtain an almost pure population of transgenic HSPCs prior to transplantation into the mice. This significantly increased the frequency of human reporter leukocytes in the HR4lox-mice. This was an important protocol optimization in light of the model being used for studying latency, where the maximum readout possible is needed because of the rarity of latently infected cells. Indeed any cell that does not carry the reporter transgene that becomes infected is an undetectable event. Unexpectedly, HSPCs sorted by FACS gave rise to effectively reconstitution hu-mice in an all-or-nothing manner. As this phenomenon was not observed with HSPCs that had only been transduced, it is conceivable that the shearing stresses applied to the HSPCs during FACS may affect their activation state or viability, or perhaps induce their differentiation, such that the cell population transplanted in the newborn mice are no longer

able to sustain hematopoiesis. Still, as generating HR4lox-mice with a pure population of transgenic HSPCs was beneficial to the model, we accepted the cost of this optimization.

Productive infection experiments, revealed that GFP+RFP+ could readily be isolated from spleen, bone marrow and peripheral blood samples of HOCC-infected reporter mice and thereafter shown to produce p24 in co-culture assays (VOA) and also shown to harbor viral DNA (HIV *gag* PCR). These proof-of-principle experiments validated that it is possible to identify and purify infected cells from the HIV reporter mouse model based on RFP expression (Figure 5). As the expression of Cre in HOCC is driven by the constitutive CMV promoter, theoretically integration of viral cDNA is not necessary for Cre to be expressed [50, 229]. This means that in the model, RFP expression marks cells in which at least viral entry, reverse transcription and nuclear translocation of the viral genome have effectively occurred. In other words, cells containing integrated provirus as well as unintegrated forms of HIV, known as LTR circles, which do not lead to productive infections [50], can be identified by proxy of RFP. However, we believe this limitation is minor as the efficacy of transcription and translation from unintegrated forms of HIV is low [50, 229] and generally confined to certain viral genes (i.e. Nef and Tat [230]), thus the overall contribution of these viral forms remains questionable. Generally the highest frequency of infected (i.e. RFP+) cells were found in the spleen, which is expected as the spleen of NSG hu-mice has a high frequency and abundance of CD4+ T lymphocytes [157] and HIV replication primarily occurs in lymphoid tissues [51, 90, 92]. Infected cells occurred at a frequency on the order of 10^4 infectious units per million (IUPM) GFP+ CD4+ T cells and principally displayed an activated (HLA-DR+), memory (CD45RA-) phenotype (Figure 4), as has been widely observed in patients or tissues of SIV-infected macaques [89, 90, 213, 220, 231]. Furthermore, work done in our lab by Dr. Mary-Aude Rochat using YU-2 infected NSG hu-mice established that at one month post infection, on average the spleen contained 21647 (± 29032) copies of integrated HIV DNA per million human (CD45+) cells. The comparability of frequencies of RFP+ cells in our model of productive HIV infection with those from Mary-Aude's study provide a certain measure of confidence that our model correctly recapitulates HIV infection in hu-mice, whilst capturing the pathophysiology observed in HIV-infected individuals. Nevertheless the intensity of the RFP signal as measured by flow cytometry was on average low, with the positive population just bulging out from the RFP negative population, rather than separating distinctly (as seen *in vitro* in primary HR4lox-CD34+ cells infected with HOCC stock, Figure 2G). This is likely a consequence of the CMV promoter used to drive RFP expression in infected cells, as despite it being a strong constitutive viral promoter, some studies describe only low levels of gene expression from this promoter in primary T lymphocytes and CD34+ cells [232-234], as well as tissue-specific promoter silencing in both immunocompetent and immunodeficient mice [235, 236], possibly due to methylation [237]. The consequence of having both a low frequency of RFP+ cells and a weak RFP signal intensity in infected cells, is the difficulty in isolating these cells, which means a) negative controls must always be included for setting strict

sorting gates; and b) tissue samples from a number of animals need to be pooled to have enough cells.

In the first proof of principle experiments, we had tried to include an anti-p24 antibody in the staining panel for flow cytometry analysis, as this would have provided a direct means for assessing whether RFP+ cells truly are productively infected. Intact infected cells harbor p24 intracellularly, so staining for p24 requires fixation and permeabilization of the cells prior to antibody labeling. In addition, the only available fluorochrome-conjugated anti-p24 antibodies available are proprietary products from Beckmann Coulter and the company has only come out with FITC- and PE-conjugated anti-p24 antibodies, both of which are not suitable with our model as their emission spectra are too similar to those of GFP and RFP, respectively. Thus, in addition to the numerous washing steps involved in intracellular staining, we had to use a secondary fluorochrome-conjugated antibody to label a non-conjugated primary p24 antibody. After intracellular p24 staining of splenocytes from viremic HR4lox-mice, we failed to clearly detect any RFP+ events by flow cytometry, probably because there were too few cells for such a low-yield staining procedure compounded by the inherently low frequency of infected cells *in vivo* at any given time [89, 213]. Besides, intracellular staining is intrinsically limited, as certain downstream analyses, such as VOA, cannot be performed because of the fixation and permeabilization of samples. Therefore, in this study we relied on both viral cultivation (VOA) and detection of cell-associated viral DNA (HIV-Gag PCR) to assess viral burden in different tissue compartments. To quantify p24 levels in the VOA culture supernatant we used a relatively novel ultra-sensitive p24 single-molecule digital ELISA (SIMOA, Quanterix) [119]. The p24 digital ELISA provides comparable results to HIV nucleic acid amplification testing [238] and has an LOD of 0.0025 pg/ml and LOQ of 0.0076 pg/ml, however it is noteworthy that one of the thirteen HIV-negative samples tested produced a reading of 0.0037pg/ml [119]. Therefore, a key aspect to accepting the veracity of very low p24 readings is seeing a persistence, and ideally growth, of the p24 signal over time. Owing to the p24 digital ELISA technology, Passaes C. P. et al demonstrated that a single productively infected CD4+ T cell produced between 0.096 and 0.36 pg/ml of p24 and released 0.29 pg/ml (median) of p24 into the culture supernatant after overnight co-culture with autologous activated CD4+ T cells [224]. Exponentially increasing p24 levels are expected in cultures where virus outgrowth from infected cells occurs and a propagating infection is successfully established. We found that VOA on cells isolated from productively infected HR4lox-mice resulted in a broad range of p24 levels in the culture supernatants (0.019 pg/ml to saturation, i.e. >200 pg/ml); with a trend towards higher p24 levels in the VOA cultures started with large amounts of cells (n > 30000) (Figure 5). Considering that the PBMCs of acutely HIV-infected patients principally harbor defective proviruses [47], this observation seems logical; i.e. the likelihood of having at least one cell carrying replication-competent provirus capable of establishing a spreading infection in the culture increases with the size of the starting pool of infected cells. Importantly, the low level p24

signals observed in the VOA cultures persisted overtime, suggesting they are a result of p24 release from infected cells rather than background noise from the digital p24 ELISA. In addition, although only low levels of p24 were detected in the VOA for the GFP+RFP+ populations, in these cultures at most about 400 cells were put in culture. Importantly, besides the replication-competency of the provirus, establishment of a propagating infection (i.e. exponential viral growth) from an infected cell has been shown to be dependent on a viral RNA expression threshold (i.e. number of HIV RNA copies) [228, 239]. In fact, Hataye et al. found that the majority of cells put in viral outgrowth cultures died before release of progeny virions, and in the event of release of virions, these frequently failed to establish propagating infections in secondary cultures [228]. Thus, it is likely that a large proportion of the GFP+RFP+ cells isolated from viremic HR4lox hu-mice harbored defective proviruses, died in culture before releasing infectious viral particles, or in which viral transcription was not sufficient to engender exponential viral outgrowth, resulting in viral extinction and only low levels of p24 released from dead cells or abortive viral replication. Of note, it is highly improbably that the p24 detected in these assays came from virions attached to the surface of the isolated cells, as the Env-receptor bounds (~30 pN [240]) would not withstand the shearing forces applied on the cell surface during FACS, where a fluid pressure of 70psi (0.482 N/mm^2) was used, as well as the numerous washing steps involved in the sample preparations (pre- and post-FACS). If somehow virions did remain attached to cells, anyways we would expect the same amount of p24 on RFP- cells. Furthermore, the congruency of the HIV *gag* PCR results with the VOA results, except for CD4+GFP+RFP- PBMCs, further support that the GFP+RFP+ cells isolated from the hu-mouse model are truly infected cells. The discrepancy between the low p24 levels measured in the VOA culture supernatants and lack of detection of HIV *gag* by qPCR in the CD4+GFP+RFP- PBMCs of productively infected HR4lox-mice implies that the LOD of the p24 digital ELISA may actually be higher in our hands than reported [119], or that there are technical issues with our whole-genome amplification or HIV *gag* PCR assays. Alternatively, some underlying biological mechanism may be at hand, such as a large fraction of infected PBMCs carrying only unintegrated HIV DNA [50], which could still support Cre expression (LTR independent) and p24 expression (early gene), but would be degraded and not detectable by PCR because of the labile nature of unintegrated viral forms. Ultimately, the reason for this discrepancy is unclear and further experiments are needed to verify whether this phenomenon is repeatedly observed in PBMCs, which would point towards a biological effect, or occurs randomly, which would support a technical effect. We believe the latter is more likely. Taken together, despite the drawbacks of low RFP signal intensity and small samples sizes, these results confirm that in the context of productive infection, the HIV reporter mouse model presented herein allows for the isolation of *in vivo* infected cells based on RFP expression. By suppressing viral replication to undetectable levels in infected HR4lox-mice with cART (i.e. latency scenario), we then sought to assess whether the model was suitable to study HIV latency.

In the latency scenario, we initially found that splenic RFP+ cells arose at a frequency of ~0.2% and displayed a naïve phenotype in the HIV reporter hu-mouse model (Figure 6B-C). However, FACS resulted in a low yield of GFP+RFP+ cells and strikingly neither viral replication nor HIV DNA could be detected in the CD4+ cells. Therefore, we repeated the experiment with a new cohort of aviremic HOCC-infected HR4lox-mice, pooling samples from five animals. This last experiment revealed that ultimately, when comparable numbers of cells for experimental and control samples (i.e. aviremic HOCC- or WT HIV-infected HR4lox-mice, respectively) were analyzed and sorted by FACS, the RFP signal intensity produced by the extremely rare RFP+ cells in the experimental samples was too weak to enable clear discrimination of latently infected cells from background noise (Figure 6D). Therefore, in hindsight, the validity of the results from the first set of latency studies with the HIV reporter mouse model is very questionable and thus these results, although initially thought to be intriguing, should be disregarded. Nevertheless, in this last latency experiment, although none of the isolated GFP+RFP+ cells could establish a propagating infection in the VOA cultures within one-month of culture, p24 was uniquely detected at low levels in the culture supernatant of GFP+RFP+ CD4+ splenocytes and presumably monocyte-lineage (i.e. CD3-CD19-) bone marrow cells (7.42 and 0.589 pg/ml, respectively), when considering a revised LOQ of 0.055 pg/ml (Figure 6E-F). A revised LOQ was implemented for the p24 digital ELISA readings as the negative control wells gave rise to p24 readings reaching up to 0.055 pg/ml, implying that in our hands the p24 digital ELISA had a higher level of background noise than reported in other studies [119, 224, 238]. The positive p24 reading from the VOA of CD4+GFP+RFP+ splenocytes was measured only at the last sampling time point, so it is not possible to determine whether, with longer culture times, p24 levels would have progressively increased, indicative of viral outgrowth, or been sustained at a low level, suggestive of ineffective reactivation and abortive replication of a latent provirus. The latter scenario is a plausible explanation for the single, early p24 signal observed in the culture supernatant of the CD3-CD19-GFP+RFP+ bone marrow cells. Although these low p24 readings must be considered with caution due to the absence of longitudinal measurements of p24 in these two cultures, the p24 levels are in accord with other studies quantifying, by digital ELISA, p24 release from stimulated CD4+ T lymphocytes from aviremic HIV patients [73, 224, 241]. Specifically, in these studies, p24 levels observed in viral induction assay culture supernatants ranged from 0.017 to >100 pg/ml, with some cultures not giving rise to any p24 production, as well as instances of abortive (transient) viral replication. Importantly, the type of stimulation or LRAs used to induce latent provirus greatly affect the outcome of viral reactivation assays in aviremic HIV patients [59, 145, 242], which brings forth the question of whether the VOA conditions used herein were sufficient to effectively reactivate latent provirus [86, 114], and if so whether reactivated latently infected cells were unable to overcome a threshold in viral replication (i.e. viral RNA or p24 threshold) required for successful viral propagation [59, 68, 228, 243]. Finally, it is also feasible that a significant portion of RFP+ cells in our aviremic HOCC-infected reporter mice harbored defective provirus [47, 85, 195]. In any

case, ultimately p24 was only detected in the cultures containing GFP+RFP+ cell populations, suggesting that RFP expression in the hu-mouse model under latency conditions marks latently infected cells.

In summary, the resolution of the HIV reporter mouse model, dictated by the intensity of the RFP signal produced by infected cells, is too low to effectively identify and purify HIV reservoir cells. In addition, a major limitation in the model is the relatively low Cre-lox recombination efficiency, which drastically limits the number of infection events that can be detected. Despite these limitations, the HIV reporter hu-mouse model offers a means to specifically mark productively infected cells *in vivo*, thereby allowing these cells to be identified, characterized and isolated by FACS. Thus, the model could be useful for identifying sites of early viral replication associated with the route of infection, or to more extensively assess cell types and tissue compartments that support viral replication throughout the human immune system. Furthermore, during the process of developing and validating this model, we gained valuable insights into key features of the model that would need to be modified in order to improve it, and thereby make it suitable for studying the HIV reservoir. These are discussed in the subsequent section.

2.3.2. Further directions and optimizations

In order to improve the sensitivity and readout of the model, both the strength of the reporter gene signal and the overall level of expression of the reporter transgene in the human immune system of hu-mice need to be increased. Firstly, to address the latter without compromising human leukocyte reconstitution in the hu-mice, and to circumvent the need to use FACS to obtain pure populations of HR4lox-transduced HSPCs, a practical modification would be to use a constitutively expressed cell surface receptor in the place of GFP, so that effectively transduced HSPCs can be selected by magnetic-activated cell sorting (MACS). MACS is generally considered to be a more gentle method for sorting and the lower purity of MACS sorts compared to FACS would not be an issue in our case, as a 100% purity of transgenic HSPCs would not be absolutely necessary [244]. Thus, reporter hu-mice with good engraftment and transgene expression levels could easily be generated. In addition, such a cell surface receptor would allow for an enrichment of transgenic cells from reporter hu-mice samples prior to the planned assays (e.g. FACS), thereby helping to reduce background noise. For this purpose, an interesting option would be to use the fusion protein CherryPicker from Clontech, which is composed of the fluorescent protein mCherry and the transferrin receptor membrane-anchor domain [245], or a similar type of membrane-targeted protein [246]. Such membrane-anchored fluorescent proteins would allow the evaluation of transduction levels and identification of transgenic cells by flow cytometry as well as selection of transduced cells by MACS. Secondly, to increase the strength of the reporter gene signal, alternative promoters used to drive the expression of the reporter gene should be tested to find the

promoter that mediates the highest levels of reporter gene expression in human primary cells. In addition, it would be beneficial to use a reporter gene in place of RFP that produces a brighter fluorescence signal, such as the yellow and orange fluorescent proteins YPet and mKO, respectively. According to a quantitative assessment of fluorescent proteins by Cranfill et al., the relative brightness (i.e. product of the protein's quantum yield and extinction coefficient relative to that of eGFP- the bench-mark fluorescent protein) of YPet and mKO were 100.32 and 103.18, respectively, compared to 19.25 for mRFP and 37.52 for eGFP [247]. Using a brighter fluorescent reporter protein would enhance the separation between the positive and negative populations, thereby facilitating the purification of infected cells in our model. Of note, although the theoretical brightness of a fluorescent protein may be high, protein folding, the microenvironment and the excitation wavelength greatly influence the actual brightness of the protein's emission wavelength [247]. Therefore, several fluorescent proteins should be tested in human primary cells to evaluate which is most likely to yield a bright signal in the model. Moreover, as GFP and RFP have overlapping excitation and emission spectra, spill-over of the GFP signal into the RFP channel during flow cytometry analysis must be compensated for, which also reduces the RFP signal. This further supports the use of alternative reporter genes in the HR4lox transgene. Thirdly, to increase the readout of the model, the barrier to Cre-mediated recombination in infected cells could potentially be lowered by simplifying the transgene design. The current HR4lox transgene design requires two Cre-mediated events (i.e. requires Cre to act twice), for induction of RFP expression, which could be significantly reducing recombination efficiency in infected transgenic cells. Alternative commonly used strategies for *in vivo* inducible gene expression using the Cre-lox technology involve implementing a floxed stop codon in front or in the middle of the reporter gene [209, 248, 249], or using a unidirectional inversion strategy making use of special asymmetric loxP sites that flank the reporter gene [250]. As there is a risk of leaky reporter gene expression with the floxed stop codon design arising from translational read-through [251, 252], alternatively longer or synthetic stop sequences could be used [253]. Finally, as Cre-mediated recombination efficiency is mainly dictated by the availability of Cre in the cell, the promoter used to express Cre is critical [209]. In the current HR4lox reporter system, Cre expression is driven by the CMV promoter, which may be problematic *in vivo*, as discussed above. Thus, again, using an alternative promoter to drive Cre expression in the HOCC genome may also help improve resolution in the HIV reporter hu-mouse model.

Furthermore, we have also been exploring an alternative design for a bioluminescent HIV reporter system, which would similarly be implemented into the NSG humanized mouse model, to generate a hu-mouse model where productively HIV infected cells can be tracked in real-time by bioluminescence imaging (BLI). As we are still in the developmental stages of this second reporter model, only the concept behind the reporter system is briefly described herein. Our BLI HIV reporter system is based on using a highly Tat-dependent promoter to drive the expression of a

luciferase reporter gene. The promoter is derived from the HIV LTR, but with modifications introduced to reduce the promoter's basal activity and increase Tat-dependent transcription. Tat is expressed early after HIV infection [230, 254], thus employing a Tat-dependent promoter to drive luciferase expression permits to specifically monitor productive HIV infection in transgenic cells by BLI (after injection of the luciferase substrate, D-luciferin). An advantage of this model is the use of wild type HIV, circumventing any issues of compromised infectivity and virus fitness that can arise with genetically modified HIV clones. The major concern with this model is achieving minimal unspecific luciferase expression, resulting from basal transcription from the modified LTR or read-through transcription from upstream promoters. Thus, in order to obtain high specificity and minimize background noise in our BLI reporter model, it is imperative to have a promoter that is ideally completely Tat-dependent, as well as implement insulator sequences to minimize the effect of neighboring genes. To this end, I supervised Anja Forrer (MD Master's project) in the design and generation of modified LTR promoters with minimal basal (Tat-independent) transcriptional activity and maximal Tat-dependent activity. To generate Tat-dependent promoters, host transcription factor binding sites in HIV WT LTR, which were not vital for promoter functionality, were removed and the TAR region of the HIV LTR promoter was duplicated. Thereby Anja Forrer created a number of different modified LTRs and tested their *in vitro* transcriptional activities. In summary, we found that duplication of the TAR region was most beneficial in increasing reporter gene expression in the presence of Tat. Surprisingly, all the modified LTRs also displayed increased levels of basal reporter gene expression. Therefore, further work is required to reduce basal activity of the promoter destined to be used in our BLI HIV reporter system. Moreover, in order to maximize the sensitivity of such a BLI reporter model, I tested different luciferases for their *in vivo* bioluminescent properties. I found that a particular red-shifted, human codon optimized luciferase called PLR2 and generated by Professor Bruce Branchini (Connecticut College, Connecticut, USA) allows for much more sensitive detection of cells in deep tissues than the commonly used Luc2 gene from Promega. Thus, this new luciferase reporter gene seems promising for our BLI HIV reporter model.

In conclusion, we plan to continue working on the HR4lox reporter system to render it more suitable for the study of HIV latency *in vivo*, with the ultimate goal of isolating and characterizing the genomic and transcriptomic profiles of reservoir cells. In addition, we will pursue the development of our BLI HIV reporter hu-mouse model, building on the work done at the beginning of my PhD and by Anja Forrer. We envisage using the BLI model to longitudinally monitor by BLI in situ viral replication in the context of different infection paradigms, as well as explore the effects and efficacy of LRAs on the viral reservoir.

2.4 Materials and Methods

Ethics statement: All animal experiments were approved by the Cantonal Veterinary Office (#93/2014) and performed in accordance to local guidelines (KTSchV, Zurich) and the Swiss animal protection law (TSchG). Procurement of human cord blood was approved by the Ethical Committee of the University of Zurich (EK #EK1103) and cord blood was collected with informed written consent from donors.

Cells: MOLT-4/CCR5 cells were cultured in R-10 medium (RPMI 1640 (Sigma-Aldrich) supplemented with 10% FBS (Biochrom), 1% 2 mM L-glutamine and 1% penicillin-streptomycin (both from Life Technologies)). 293T cells were cultured in D-10 medium (DMEM (Sigma-Aldrich) supplemented in the same way as the R-10 medium). PBMCs were isolated from buffy coats, obtained from the local blood donation center in Zurich, by Ficoll (Alere Technologies AS) gradient centrifugation. CD4+ T cells were purified from freshly isolated PBMCs using anti-CD4 MicroBeads (Miltenyi Biotec) with a purity ≥90%. Cells were stored frozen in liquid nitrogen until use.

Wild-type HIV production and titration: Wild-type NL4-3 or YU-2 viruses were produced by transient transfection of 293T cells with polyethylenimin (Sigma-Aldrich) and the pNL4-3 (NIH Aids reagent program, Cat. No. 114) or pYU-2 (NIH Aids reagent program, Cat. No. 1350) plasmid, respectively. After 3 to 18 hours incubation, the supernatant was exchanged for fresh D-10 medium. Forty-eight hours post-transfection, the supernatant was harvested, filtered (0.22um, Steriflip, EMD Millipore), aliquoted and frozen at -80°C until use. Virus stocks were titrated as described in [255]. Briefly, CD8-depleted PBMCs from three donors were activated with recombinant IL-2, PHA and OKT3 antibody for two days and then challenged with serial dilutions of virus stock in order to determine the TCID50 by p24 ELISA.

HR4lox lentiviral particle and HOCC production and titrations: HR4lox lentiviral particles were produced (by Dr Petitjean) by calcium phosphate co-transfection of 293T and equal amounts of the HR4lox lentiviral transfer plasmid, pCMV-delta-R8.2 packaging plasmid (Addgene plasmid # 12263) and pMD2.2-VSVG envelope plasmid (Addgene plasmid # 12259). HOCC virus was produced (by Dr Petitjean) by calcium phosphate transfection of 293T cells with the HOCC plasmid (pBR-NL4-3 Nef-codon optimized-CMV-CRE). For both preparations, 12-24 hours post-transfection the cells were washed and fresh D-10 medium without FBS was added. After 72 hours the supernatant was harvested, filtered and concentrated on a sucrose cushion (20% sucrose in PBS Ca2+ Mg2+). The virus particles were then resuspended in PBS Ca2+ Mg2+ for 4 hours at 4°C, before being aliquoted and stored at -80°C until use. HOCC virus stocks were titrated by challenging $5x10^4$ 2-day activated (0.5ug/ml PHA, 20U/mL IL-2) PBMCs from healthy donors with serial dilutions of virus stock. After 5 days, cells were stained for intracellular Cre (primary mouse anti-Cre recombinase antibody (Abcam) plus secondary anti-mouse APC antibody) and anaylsed

by flow cytometry to estimate the frequency of Cre+ cells and thereby calculate the infectious units per mililiter (IU/mL). The TCID50/ml for HOCC virus stocks was then estimated from this IU/mL titer by dividing the latter by 0.7 (based on the Poisson distribution and [256]).

Isolation of CD34+ cells from umbilical cord blood: CD34+ HSPCs were isolated from human cord blood by Ficoll (Alere Technolgies AS) gradient centrifugation and with anti-CD34 immunomagnetic beads (Miltenyi Biotec). HSPCs and the flow-through (i.e. non-target) fractions were either used directly or stored frozen in liquid nitrogen until use.

Transduction of umbilical cord blood-derived CD34+ HSPCs: Freshly isolated human HSPCs (see above) were cultured in CellGro (CellGenix) supplemented with SCF 20 ng/mL, FLT3 20 ng/mL, TPO1 20 ng/mL, IL-3 20 ng/mL (all cytokines from Immunotools) at a concentration of 10^6 cells/mL. Eighteen to 24 hours after activation, LentiBlast-B (Oz Biosciences SAS) at a concentration of 1:1000 and HR4lox lentiviral particles were added to the cells for approximately 48 hours. Transduction efficiency was assessed based on GFP expression by flow cytometry and whenever possible, cells were directly transplanted into newborn NSG mice (see below), or otherwise stored frozen in liquid nitrogen until use. In later experiments, prior to transplantation, the transduced HSPCs were subjected to FACS (FACS Aria III 5L, BD Biosciences) to obtain a pure population of GFP+ (transduced) HSPCs.

Generation, infection and treatment of hu-mice: Hu-mice were generated as described in [163]. Briefly, newborn immunodeficient NOD-SCID-IL2Rgamma$^{-/-}$ (NSG) mice (Jackson laboratory) were sub-lethally irradiated (1 Gy) and transplanted intrahepatically with approximately $1.0\pm0.5\times10^5$ untransduced or transduced CD34+ HSPCs. Mice were bred and maintained in specific pathogen free conditions (Biologisches Zentrallabor at the University Hospital Zurich) in individual ventilated cages and were given autoclaved food and water. Mice were monitored once weekly according to a standard score sheet. Twelve to 16 weeks after transplantation, human engraftment and *de novo* human immune system reconstitution in the mice were assessed by flow cytometry (CyAn™ ADP Analyzer, Beckman Coulter) by staining PBMCs with anti-human monoclonal antibodies against the following cell-surface markers: CD45-Krome Orange (Beckman Coulter), CD3-PE, CD4-Pe Cy7, CD8-BV421, CD19-APC (all from Biolegend).

Hu-mice were infected intraperitoneally with HOCC, NL4-3 or YU-2 at a dose of $1.0\text{-}1.2\text{x}10^6$ TCID50 per mouse, unless stated otherwise. For the latency studies, infected hu-mice were treated with Abacavir (ABC, 100 mg/kg per day), Lamivudine (3TC, 150 mg/kg per day) and Dolutegravir (DTG, 15 mg/kg per day) (all from ViiV Healthcare) dissolved in a sucralose solution replacing the normal drinking water as described in [24]. Briefly, drug tablets were manually pulverized, weighed to get the desired amount of active compound (assuming an average body

weight of 20g per mouse and consumption of 5mL per day), added to MediDrop® sucralose solution (ClearH2O) and mixed for at least 15 minutes using a magnetic stirrer (protocol kindly provided by Julien Villaudy). The cART solution was then given in place of the normal drinking water and exchanged twice a week to avoid issues with the drugs precipitating.

Quantification of viremia: Viremia in the hu-mice was quantified using an in-house viral load assay implemented in our lab by Dr. Audigé based on the protocol described in [257]. RNA was extracted from plasma using the QIAamp® Viral RNA kit (Qiagen), eluted in 30uL AVE buffer and reverse transcribed using the iScript™ cDNA Synthesis Kit (Bio-Rad). A HIV-1 gag qPCR was then performed on the cDNA in a 50uL reaction containing 1X Maxima Hot Start PCR Master Mix (Thermo Scientific), 1 uM forward primer (5'-CAAGCAGCCATGCAAATGTTAAAAGA-3' [257]), 1 uM reverse primer (5'-TACTAGTAGTTCCTGCTATGTCACTTCC-3' [258]) and 0.3 uM probe (5'-FAM-TGCAGCTTCCTCATTGATGGT-BHQ-1-3' [259]). Experiments were performed in duplicates on a IQ™5 real-time thermocycler (Bio-Rad Laboratories, Inc), with the following cycling protocol: 95°C 4 min, 50x (95°C 5 s, 55°C 5 s, 60°C 30 s). Standards were prepared fresh for each qPCR experiment by diluting filtered supernatant from PBMCs infected with NL4-3 (aliquots stored at -80°C) containing a known number of HIV RNA copies/ml (quantified by COBAS® AmpliPrep/COBAS® TaqMan HIV-1 Test, Roche), and co-extracted with the test samples. qPCR data was analysed with iQ™5 Optical System Software v2.1 (Bio-Rad Laboratories, Inc).

Flow cytometry and FACS: For all flow cytomerty and FACS experiments besides engraftment checks, acquisitions were perfomed on a LSRFortessa™ II and FACSAria™ III 4L (both from BD Biosciences). Cell suspensions were labeled with anti-human monoclonal antibodies targeting the following cell-surface markers: CD45-BV510, CD3-APC-Cy7, CD4-PerCp, HLA-DR-BV421, CCR7-PE-Cy7, CD45RA-APC (all from Sony Biotechnology Inc), and CD19-AlexaFluor-700 (in 1st panel with CD19) or CD19-BV711 (in 2nd panel with CD19) (both from BioLegend). PBS or FACS buffer (PBS containing 2% FBS and 0.05% sodium azide) were used for washing and reagent dilutions. Before acquisition cells were fixed with 1% Paraformaldehyde (Sigma-Aldrich). Data were analyzed with FlowJo 10 software (FlowJo LLC).

Detecting Cre in circulating (plasma) viruses: The amount of Cre recombinase present in the plasma of HOCC-infected hu-mice was quantified by RT-qPCR. RNA was extracted from plasma using the QIAamp® Viral RNA kit (Qiagen), reverse transcribed and the cDNA subjected to a Cre-qPCR (assay established by Dr Petitjean in Prof. Benkirane's laboratory).

Viral outgrowth assay on sorted cells and quantification of p24 antigen in culture supernatants: Cells isolated by FACS from infected (viremic or aviremic) HR4lox hu-mice were washed in PBS and then co-cultured with permissive cells under T-cell stimulatory conditions to

induce viral outgrowth, which was assessed by longitudinally quantifying p24 antigen concentrations in the culture supernatants. Two different VOAs were used in this study. The in the first VOA, used in all experiment except the last latency study, sorted cells were co-cultured with 1 million CD4+ T cells from healthy donors in R-10 supplemented with 100U/mL human rIL-2 (NIH Aids Reagents Program Cat. No. 136) and 0.5ug/mL PHA (Sigma-Aldrich). After 72 hours, supernatants were collected, the remaining culture medium was exchanged to remove the PHA and 1 million fresh 3-day activated (rIL-2 and PHA) CD4+ T cells from the same donor were added. Thereafter, supernatants were collected every 3-4 days for 3 weeks. At day 9, 1 million fresh 3-day activated (rIL-2 and PHA) CD4+ T cells from the same donor were again added. The second VOA was conducted as described in [260], except that sorted cells were not cultured in limiting dilutions (too few cells). Briefly, sorted cells were cultured in Super T cell medium (STCM; containing 100U/mL human rIL-2 and 1% T cell growth factor) on irradiated PBMCs from healthy donors and 1ug/mL PHA. After 24 hours, the medium was exchanged to remove the PHA and MOLT-4/CCR5 cells were added to help propagate reactivated virus. Five days later supernatants were collected and the cells split into two, seeding the second half of cells into a new plate, and fresh STCM added. For 4 weeks, supernatants were collected, cultures split in two and fresh STCM added every 4-7 days. For both VOAs, p24 antigen concentrations in the culture supernatants were quantified using the Simoa™ p24 Antigen Assay (Quanterix) (ultrasensitive digital immunoassay, LOD = 0.0133pg/mL) as described in [224].

Whole genome amplification (WGA) and detection of HIV DNA in sorted cells: WGA was done by multiple displacement amplification (MDA) using the illustra™ Single Cell GenomiPhi™ DNA Amplification Kit (GE Healthcare) as specified in the kit manual, except for the following optimizations: 2-10uL of cell suspension in PBS were used as starting material; lysis and neutralization buffer volumes were scaled up to 2uL (doubled if > 1000 cells were in the starting cell suspension). DNA quantification of the MDA products was done in duplicates with 1uL of amplified DNA using the Quant-iT™ PicoGreen™ dsDNA Assay Kit (Invitrogen). The presence of HIV DNA (gag) in the MDA products was assessed using our in-house HIV-1 gag qPCR assay (described above), carried out in duplicates on 1uL of amplified DNA. The remainder of the MDA products were stored at -80°C.

Statistical analysis : Statistical analyses were performed using GraphPad Prism 7 (GraphPad Software). D'Agostino-Pearson omnibus normality test was always applied to data sets to first assess whether parametric tests could be used. Accordingly, data were then subjected either to t-tests or ANOVAs (or non-parametric equivalents). Significance was considered when p-values were ≤0.05

3. CHAPTER 3

Dissecting the contribution of APOBEC3G to HIV diversification and adaptation to antiviral monotherapy in humanized mice

Disclaimer and author contributions

The work presented in this chapter is the fruit of a long-term collaboration with Professor Simon's group, extending over more than eight years and involving the following persons:

Audrey Fahrny[1], Matthew M. Hernandez[2], Anitha Jayaprakash[3], Marsha Dillon-White[3], Gustavo Gers-Huber[1], Annette Audigé[1], Ravi Sachidanandam[3], Viviana Simon[2,4,5], Roberto F. Speck[1]

1. *Division of Infectious Diseases and Hospital Epidemiology, University Hospital of Zurich, University of Zurich, Zurich, Switzerland*
2. *Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY, USA*
3. *Department of Oncological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA*
4. *Global Health and Emerging Pathogens Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA*
5. *Division of Infectious Diseases, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA*

As we are still in the process of writing the manuscript for publication, but have not yet submitted, this chapter consists of a draft of our manuscript. The introduction and discussion sections of this chapter were written by myself, whereas the methods and results sections were written together with M.M.H. All the *in vivo* experiments were performed by Professor Speck's group (A.F, G.G.H, A.A) and all the sequencing work was performed by Professor Simon's group (M.M.H., A.J., M.D.W). Data were analyzed by A.F., M.M.H., and R.S. Over the past four years, I have had the lead on this project, planning and running experiments, analyzing the data I generated, as well as brainstorming, discussing and interpreting all results together with M.M.H. M.M.H and I will share first co-authorship on the manuscript. Note this study is also presented in M.M.H.'s MD PhD thesis.

3.1. Introduction

Genetic diversification is key feature of HIV and essential for viral evolution, adaptation and escape from immune surveillance and antiretroviral therapy (ART) [173, 261]. Accordingly, viral diversification in HIV-infected patients is associated progressive HIV disease and underlies the rapid emergence of drug resistant HIV strains. HIV's extensive sequence variation is widely attributed to the low fidelity of the viral reverse transcriptase (RT), which induces an estimated $3x10^{-5}$ mutations/base pairs/cycle [20, 174, 175], HIV's high replication rates (10^{10} – 10^{11} virions/day [21, 176]) and recombination between co-packaged genomes [22, 23]. Another proposed driver of viral diversification is sublethal mutagenesis of the HIV genome induced by host deaminases from the APOBEC3 (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3) family. The APOBEC3 family consists of seven cytidine deaminases (termed A3A to A3H), which contribute broadly to human innate antiviral immunity and restrict HIV to varying degrees [71, 180-183]. A3G is amongst the most potent anti-HIV APOBEC3 enzymes, but A3D, A3F and A3H also display anti-HIV activity [71, 183]. A3G is incorporated into budding virions and therefore exerts its antiviral activity in the next round of infection (i.e. in cells newly infected with virions containing A3G). It restricts HIV replication by deaminating cytidine residues in the nascent HIV cDNA to uridine residues during reverse transcription, thereby introducing a high number of G-to-A mutations in the proviral genome, which generally result in defective proviruses [184]. In turn, HIV evolved to evade APOBEC3-mediated restriction by encoding the viral infectivity factor (Vif). Vif is a HIV accessory protein that efficiently inhibits encapsidation of APOBEC3 enzymes into progeny virions, mainly by inducing their degradation through the host ubiquitin-proteasome pathway [185, 186]. Notably, although Vif is essential for HIV propagation *in vivo* [262-264], HIV variants encoding *vif* alleles with only partial APOBEC3 neutralizing activities have been identified and recovered from HIV infected patients [187-189]. Moreover, partially defective Vif viruses in patients have been associated with viral diversification [265], ART-failure [189, 190] and co-receptor usage evolution [266]. These findings suggest that complete APOBEC3 neutralization is not necessary for HIV survival as a population *in vivo* and that Vif polymorphisms enhance HIV's evolutionary potential and adaptive capacity. However, the role of partially defective Vif alleles and APOBEC3 activity on HIV diversification and adaptation remains highly debated. Various studies have shown that A3G diversifies HIV quasispecies using *in vitro* and *in vivo* models [179, 192, 265, 267-269] or by analyzing viral repertoires in patients [188]. However, other investigations uphold that A3G does not contribute to HIV's evolutionary potential because its mutagenic effect on the viral genome is either lethal [193, 194], or insignificant compared to HIV's own mechanisms of diversification [176, 270]. Finding a clear answer to this controversial issue is complicated by the difficulty in unequivocally distinguishing sequence variation resulting from RT-error versus APOBEC3-activity. Specifically, although A3G and A3D/A3F introduce G-to-A mutations in the HIV genome within

preferred dinucleotides contexts (GG and GA, respectively) [184, 265, 271], RT also frequently introduces G-to-A substitutions, with a bias for GA targets [20, 272-274].

G-to-A mutagenesis can be both deleterious and beneficial for HIV, as it can give rise to aberrant stop codons and defective proviruses, but can also result in mutated viruses that have a fitness advantage, for example, through acquisition of mutations conferring drug-resistance. For instance, resistance to the nucleoside reverse transcriptase inhibitor, lamivudine (3TC), results from a missense mutation in the HIV *RT* gene at the methionine residue 184 (M184), with isoleucine (M184I) or valine (M184V) being the two most prevalent drug resistance codons [275-278]. As the M184I variant results from a G-to-A mutation in A3G's preferred dinucleotide context (i.e. GG-to-AG), 3TC resistance is particularly susceptible to A3G's mutagenic activity. Indeed, our collaborators showed that *in vitro*, HIV can exploit AG3 to escape from 3TC [179]. Specifically, the E45G-Vif HIV mutant (HIV-45G), which displays suboptimal A3G neutralization activity *in vitro* compared to wild type (WT)-Vif HIV, efficiently replicated in human peripheral blood mononuclear cells (PBMCs) and resulted in a diversified proviral population in which 3TC drug resistance-associated mutations were found with high frequency. These results suggest that partially active Vif alleles result in increased proviral diversity, which in turn can shape the phenotype of circulating viruses. Therefore, we hypothesized that suboptimal A3G neutralization *in vivo* would result in a highly diversified genetic viral pool, with low levels of sublethal mutagenesis in at least some viral progeny, thereby conferring a high adaptive capacity to HIV as a population (Figure 1A).

To dissect the contribution of A3G to viral fitness, diversification, and evolution dynamics *in vivo*, we used the humanized NOD-*scid* IL2Rgamma^{null} (NSG) mouse model and generated isogenic HIV clones encoding *vif* alleles with different A3G-neutralization capacities (HIV-WT: 95%; HIV-45G: 20%; HIV-SLQ: 0%). The well-established humanized NSG mouse model [158, 279] supports disseminated HIV replication and recapitulates key features of HIV pathogenesis [160, 163, 168], and thus provided us with a convenient and pertinent model to study the A3G-Vif axis within the complexity of the human immune system. Understanding the contribution of A3G to viral sequence diversification is important for the management of ART-failure and development of new therapies targeting the APOBEC-Vif axis. In this study, we found that the partially defective Vif mutant HIV (HIV-45G) displayed a superior adaptation phenotype in the context of 3TC monotherapy compared to WT HIV, despite inferior fitness in natural infections. We show that this phenotype is supported by high nucleotide diversity in proviral quasispecies landscapes, characterized by an A3G-mutational footprint, and a preferential emergence of M184I resistant variant, supporting the notion that sublethal A3G activity can enhance the evolutionary and adaptive potential of HIV.

3.2. Results

3.2.1. HIV mutant with partially defective Vif can still replicate *in vitro*

To address the role of A3G in shaping viral diversification and evolution, we generated isogenic Vif mutant HIV clones with variable activities against APOBEC3G *in vitro*. These HIV clones are derived from the HIV-1 NL4-3 consensus sequence and differ from one another only by a few nucleotides in *vif*: wild type Vif (HIV-WT), E45G Vif (HIV-45G) and 144AAA Vif (HIV-ΔSLQ) (Figure 1A-B). The E45G Vif allele was initially identified in primary viral isolates [188] and differs from wild type Vif by a nonsynonymous mutation in the A3G-binding domain of Vif at residue 45 (i.e. glutamic acid substituted for a glycine). This residue has not been specifically associated with other Vif functions and is not involved in any of the other viral gene products [280, 281]. The APOBEC3 neutralization capabilities of the HIV clones used herein have previously been characterized using cell-based assays [188, 282, 283]. Whereas wild type (NL4-3) Vif displays 95% A3G-neutralization activity, the E45G Vif mutant demonstrates only partial (20%) activity against A3G and both Vif variants can equally neutralize A3F [188, 189]. HIV-ΔSLQ carries inactivating alanine substitutions in the functional BC-box motif of Vif, essential for targeting APOBEC3 enzymes for proteasomal degradation, and therefore lacks any activity against either deaminase [280, 282]. To test the ability of each virus to replicate in the presence or absence of A3G *in vitro*, we infected MT4 cells, which do not express A3G [284], and mitogen-activated, primary mononuclear cells (MNCs) derived from healthy-donor cord blood, with our three HIV clones. As expected, in the absence of A3G (i.e. in MT4 cells), all three viruses replicated to similar levels (Figure 1C). However, in primary MNCs, which endogenously express A3G, only HIV-WT and HIV-45G replicated efficiently and to comparable levels (Figure 1D).

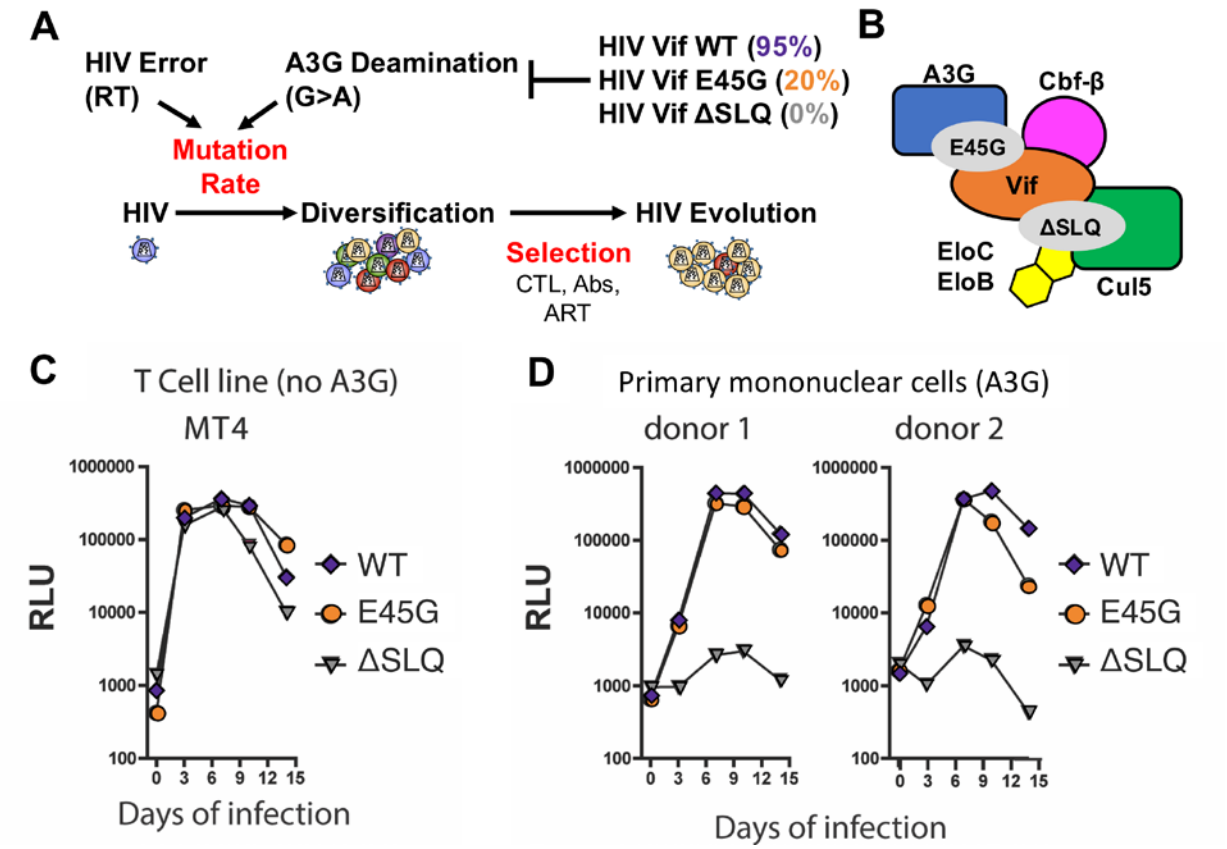


Figure 1. ***HIV-Vif variants modulate the antiviral activity of APOBEC3G in vitro.*** A. *Mutations allow HIV to diversify into a pool of genetically distinguishable quasispecies, enabling adaptation to specific pressures. Mutagenesis can result from the low-fidelity RT enzyme or A3G-induced mutagenesis.* B. *HIV-Vif (NL4-3) variants E45G and ΔSLQ display incomplete neutralization of A3G due to interference with A3G-binding and assembly of the E3-Ub-ligase complex, respectively.* C-D. *Replication of HIV-Vif variants in the MT4 T-cell line (C) and primary mononuclear cells (D).*

3.2.2. Suboptimal neutralization of A3G results in lower viral fitness in hu-mice

To assess the impact of different degrees of A3G neutralization on HIV replication *in vivo*, we infected humanized mice (hu-mice) with the three HIV clones described above and sampled peripheral blood on a bi-weekly basis from 30 up to 85 days post-infection (dpi) (Figure 2A-B and Figure 3). This allowed us to monitor plasma viremia, as well as CD4+ and CD8+ T cell population dynamics over the course of infection. Of note, our hu-mouse model recapitulates endogenous expression patterns of human APOBEC3 enzymes, with CD4+ T cells from the spleens of hu-mice expressing comparable levels of A3D, A3F, A3G and A3H as observed in human peripheral blood CD4+ T cells [166]. As A3H is resistant to NL4-3 Vif [285], to negate any confounding antiviral effects induced by A3H in our model, all hu-mice herein were reconstituted with HSPCs from donors with inactive A3H haplotypes (e.g. haplotypes I, III, IV and VI [284, 286, 287]).

Mirroring our *in vitro* results and published data [280, 282], HIV-ΔSLQ was unable to establish a productive infection in the hu-mice (Figure 2B), confirming that APOBEC3s are effectively expressed and active in our hu-mouse model in absence of Vif. On the other hand, HIV-WT and HIV-45G displayed robust viral replication at 30 dpi (Figure 2B), referred to as baseline herein, and infection rates were comparable between the two viruses (Fisher's exact test: p=0.0994). However, despite comparable (p=0.1673) baseline plasma viral loads in both groups (HIV-WT: 3.97±5.94 $x10^4$ copies/mL; HIV-45G: 8.94±10.3 $x10^4$ copies/mL), over time HIV-45G viral replication was attenuated compared to HIV-WT. Specifically, beyond 44 dpi, viremia progressively declined in the HIV-45G infected hu-mice, whereas viremia rose or remained stable in the HIV-WT infected hu-mice (Figure 3A-B). This corresponded to an average reduction in viremia from baseline to endpoint (i.e. 58, 75 or 85 dpi) of 0.77 log in the HIV-45G group compared to an average increase in viremia of 0.43 log in the HIV-WT group (p=0.0039). Taken together, these results suggest that HIV-45G exerts sufficient anti-A3G activity to propagate *in vivo*, but less effectively than HIV-WT. Furthermore, longitudinal flow cytometric analysis of PBMC populations revealed that there were no significant differences in the change in CD4+ T cells levels and CD4+/CD8+ T cell ratio over time between hu-mice infected with HIV-WT or HIV-45G (Figure S1). Monitoring CD4+ T cell levels and the CD4+/CD8+ T cell ratio provides insight about the pathogenicity of HIV, thus these results indicate that despite different replication kinetics the viruses were similarly pathogenic.

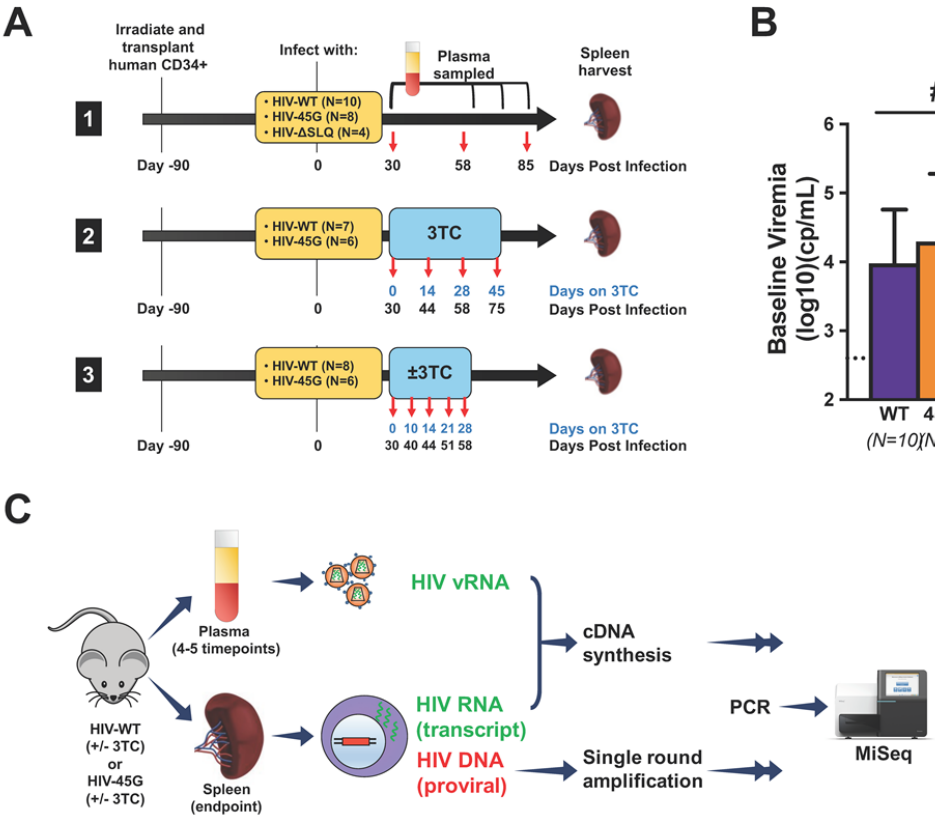


Figure 2*. In vivo experimental design and approach for sequencing HIV quasispecies in plasma and splenocytes.* A. *Design of three independent experiments to study the replication of HIV-WT, HIV-45G and HIV-ΔSLQ in humanized mice without 3TC (Exp. 1, 3) and with 3TC treatment (Exp. 2, 3)*. B. *Viremia at 30 days post infection (i.e. baseline viremia) in HIV-WT, HIV-45G and HIV-ΔSLQ infected mice across the three experiments. The dotted line represents the limit of detection of the viral load assay (400 cp/mL).* C. *Schematic depicting approach to sequence HIV quasispecies in plasma and splenocytes*.

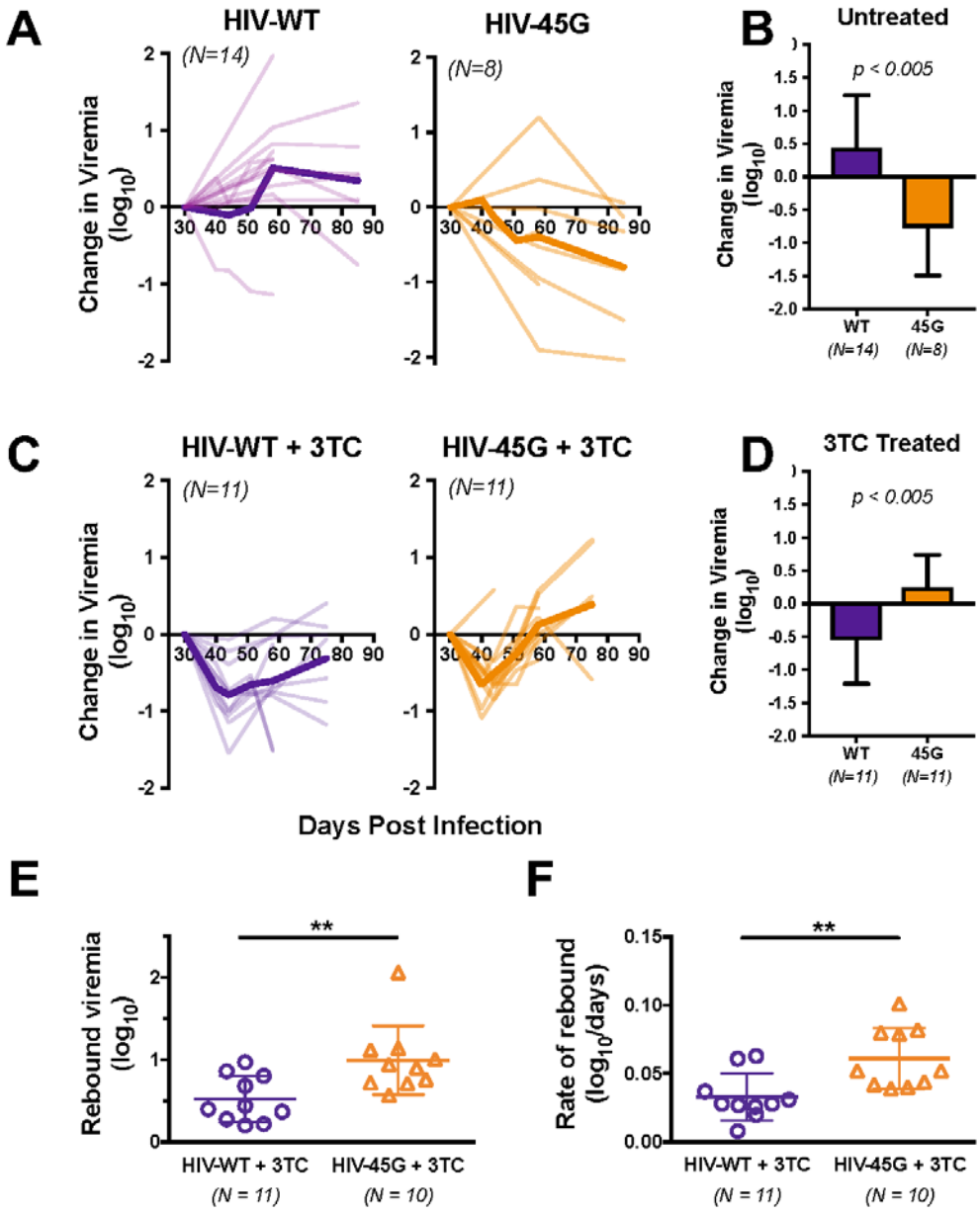


Figure 3. ***HIV-WT and HIV-45G replication in hu-mice in the absence or presence of 3TC.*** A. & C. *Change in viremia relative to baseline over time in untreated (A) and treated (C) hu-mice.* B. & D. *Overall change in viremia relative to baseline at the end of infection (mean±SD) in untreated (B) and treated (D) hu-mice*. E.-F. *Rebound viremia (E) and rate of rebound (F) in treated hu-mice. Pale lines represent individual mice and bold lines represent group means.*

3.2.3. Under selective pressure by 3TC, suboptimal neutralization of A3G is associated with superior viral fitness in hu-mice

To elucidate how a partially active Vif can exploit A3G-mediated mutagenesis to escape from antiretroviral pressure *in vivo*, we infected hu-mice with either HIV-WT or HIV-45G and 30 dpi we treated the hu-mice with 3TC for 4-6 weeks (Figure 2A). 3TC was used to apply a specific selective pressure on the viruses and was chosen because 3TC-resistance is well characterized [275-277]. Over the first two weeks of 3TC monotherapy, as expected, we observed a phase of viral restriction characterized by a decrease in viremia in 11/11 HIV-WT and 10/11 HIV-45G infected hu-mice (Figure 3C). Of note, the lone HIV-45G infected hu-mouse (#120) that did not respond to 3TC treatment, showed signs of wasting and had to be euthanized at 44 dpi and was thus excluded from further analyses. The average initial decline in viremia was similar for both HIV-WT and HIV-45G infected mice (p=0.2462), with overall an average reduction in viremia of 0.7±0.34 log and ≥ 72% of hu-mice in both groups displaying a decrease in viremia greater than 0.5 log. Virologic nadirs (i.e. lowest measured plasma viral load) were also comparable (p=0.4848) between treated hu-mice infected with HIV-WT (1.71±1.84 $x10^4$ copies/mL) and HIV-45G (2.47±2.97 copies/mL). These results indicate that both viruses were initially similarly sensitive to 3TC monotherapy. However, viral escape rapidly ensued with a rebound in viremia in 8/11 HIV-WT and 10/10 HIV-45G infected hu-mice. Nevertheless, at endpoint, average HIV-45G viremia surpassed initial baseline (i.e. pre-treatment) levels, whereas average HIV-WT viremia had decreased relative to baseline (p=0.0045) (Figure 3D). Moreover, the average amplitude and rate of rebound in viremia (from virologic nadir) were significantly greater in the HIV-45G group compared to HIV-WT group (p=0.0068 and p=0.0052, respectively) (Figure 3E-F). These findings indicate that HIV-45G was fitter than HIV-WT under selective pressure by 3TC, suggesting that suboptimal A3G-neutralization increases HIV's ability to diversify and adapt to a selective pressure.

HIV-WT + 3TC						
Mouse	**BL VL (log)**	**Nadir VL (log)**	**Log Decrease (BL to nadir)**	**BL to nadir (dpi)**	**Log Increase (nadir to rebound)[a]**	**Nadir to Rebound (dpi)**
#2321	4.109	2.602	1.509	30-58	0	58-58
#2284	5.107	4.196	1.507	30-58	0	58-58
#118	5.494	4.319	1.175	30-75	0	75-75
#117	4.858	4.787	0.071	30-44	0.276	44-58
#2320	5.318	4.266	1.052	30-44	0.371	44-58
#126	5.345	4.196	1.149	30-44	0.399	44-58
#136	4.785	4.564	0.221	30-44	0.216	44-75
#2286	5.140	4.152	0.988	30-44	0.858	44-58
#116	3.350	2.602	0.748	30-44/58	0.681	44/58-75
#125	3.942	3.549	0.393	30-44	0.795	44-75
#139	4.340	2.792	1.547	30-44	0.974	44-75
HIV-45G + 3TC						
Mouse	**BL VL (log)**	**Nadir VL (log)**	**Log Decrease (BL to nadir)**	**BL to nadir (dpi)**	**Log Increase (nadir to rebound)[a]**	**Nadir to Rebound (dpi)**
#129	5.035	4.446	0.589	30-75	0	75-75
#2282	5.467	4.923	0.544	30-40	0.903	40-51
#130	5.489	4.858	0.631	30-44	0.709	44-58
#2256	5.111	4.416	0.695	30-44	0.724	44-58
#2288	4.597	4.012	0.585	30-44	1.119	44-58
#2322	5.269	4.175	1.094	30-40	0.757	40-58
#2319	4.397	3.418	0.979	30-40	0.941	40-58
#122	4.338	3.760	0.578	30-44	1.008	44-75
#119	3.848	3.193	0.654	30-44	1.145	44-75
#141	4.083	3.236	0.847	30-44	2.061	44-75

Table 1. ***Baseline, nadir and rebound viral loads in treated hu-mice.*** *BL= baseline, VL= viral load, dpi = days post infection. [a] Rebound is defined by the highest viral load after nadir.*

3.2.4. 3TC-treated hu-mice display increasing sequence diversity in *RT* and rapid emergence of 3TC resistance mutations in plasma over time

As HIV-45G displayed important differences in replication and adaptability to 3TC compared to HIV-WT in hu-mice, we sought to identify whether these phenotypes were shaped by A3G-driven mutagenesis. To this end, we identified and quantified the HIV quasispecies present in the infected hu-mice by performing deep sequencing on 250 base pairs (bp) of HIV *RT* in serial samples of plasma viral RNA (vRNA; i.e. cell-free circulating virions), as well as cell-associated HIV transcripts (HIV RNA) and proviruses (HIV DNA) in the splenocytes at endpoint (Figure 2C). To account for PCR error and amplification biases in our sequencing workflow, we included unique molecular ID tags (UMIDs) in the primers used for cDNA synthesis (see methods and Figure S2) [288], allowing us to quantitatively sequence single genome nucleic acids. We then characterized the HIV *RT* gene of the quasispecies identified in the different viral nucleic acid compartments (plasma vRNA, splenocyte-associated HIV RNA and DNA), in order to elucidate whether the viral genotypes differed between HIV-WT and HIV-45G infected hu-mice.

Firstly, to determine whether viral sequence diversity in plasma corresponded with the different fitness phenotypes, we longitudinally measured nucleotide diversity (π), mutation rate (i.e. substitutions per base pair) and fractions of G-to-A mutations in the *RT* gene of plasma vRNA quasispecies. Nucleotide diversity is a measure used to express the degree of nucleotide polymorphisms within a given population [289]. In our case, this population is the pool of quasispecies in each infected hu-mouse. In untreated hu-mice infected with either virus, nucleotide diversity was comparable and remained unchanged over time (Figure 4A-B). In line with these observations, mutagenesis was also stable throughout the course of the experiment, without any significant accumulation of mutations over time (Figure S3A, C, E). Conversely, in both treated hu-mouse groups, after two weeks of 3TC (i.e. 44pdi) nucleotide diversity was significantly higher compared to baseline levels (π at 44dpi, HIV-WT: 0.00270; HIV-45G: 0.00293), however no further increase in diversity was thereafter observed, suggesting a saturation of diversification in the circulating quasispecies. In addition, diversity was comparable between treated groups at any time point (Figure 4C-D). The dynamics of mutagenesis in *RT* of treated hu-mice mirrored this trend (Figure S3B). As the different APOBEC3 deaminases preferentially induces G-to-A mutations in distinct dinucleotide motifs (A3G: GG-to-AG; A3D/A3F: GA-to-AA) [290], we quantified the fraction of GGs and GAs that were mutated in plasma vRNA (Figure S3D, F). This revealed that both HIV-WT and HIV-45G were predominantly and comparably impacted by GG-to-AG mutations in the treated hu-mice.

As the amino acid residue 184 in HIV RT alone determines susceptibility to 3TC, we also calculated nucleotide diversity excluding mutations at the residue 184, which we refer to as non-M184 nucleotide diversity (Figure 4E-F). Non-M184 nucleotide diversity was comparable between

groups and remained stable over time. At 44 dpi, non-M184 nucleotide diversity was 0.00105 and 0.00121 for hu-mice infected with HIV-WT and HIV-45G, respectively, which means that 39% and 41% of the diversity observed in the plasma quasispecies of treated hu-mice is driven by mutagenesis at residue 184 in HIV RT. This suggests that in the context of 3TC treatment, viral diversification in the plasma is strongly driven by drug resistance development and that 3TC is imposing an evolutionary bottleneck on the viruses.

Secondly, as 3TC resistance development is particularly susceptible to A3G-mediated mutagenesis, we wanted to test whether development of resistance to 3TC was different in HIV-45G infected hu-mice compared to those infected with HIV-WT. As discussed above, the residue 184 in HIV RT determines 3TC susceptibility, with the wild type methionine (M184) rendering the virus susceptible to 3TC and resistance predominantly including the M184I and M184V mutations (Figure 5A). These mutations confer up to 1000-fold resistance to 3TC [275-277] and can both result from RT-error. But as the M184I mutation results from a G-to-A mutation in A3G's preferred dinucleotide context (i.e. ATGG-to-ATAG), it is particularly prone to arise from A3G-induced mutagenesis. Therefore, we determined the frequency of drug susceptible (M184) sequences and drug resistance mutations (M184I/V) in the plasma quasispecies of treated hu-mice over time (Figure 5B-C). Interestingly, at baseline, HIV-WT and HIV-45G infected hu-mice had comparably low levels of drug resistant variants in their plasma (HIV-WT: 1.34±2.22% M184I and 0.56±1.45% M184V; HIV-45G: 2.92±3.79% M184I and 0.18±0.31% M184V). While the baseline levels of M184I resistance mutations were higher than M184V in HIV-WT ($p=0.0450$) and HIV-45G ($p=0.0256$) infected hu-mice, these were comparable between the two groups (M184I: $p=0.1954$; M184V: $p=0.3863$). Over time, M184I and M184V levels rapidly expanded in all but one treated hu-mouse (#2321). Of note, this hu-mouse displayed a progressive loss in human engraftment in the periphery over time, which mirrored the decline in viremia to undetectable levels by 58 dpi seen in this mouse, and at sacrifice we could not detect any human cells in the spleen. Linear regression analyses of the frequencies of drug susceptible and resistant sequences over time, revealed that in HIV-45G infected hu-mice the M184I mutation emerged at 8.7-times the rate of M184V (i.e. comparison of slopes; $p=0.0035$) (Figure 5C). On the other hand, despite similar trends in emergence of resistance mutations, in HIV-WT infected hu-mice, the M184I and M184V mutations emerged at statistically similar rates (Figure 5B).

Furthermore, in order to verify the stability of the Vif genotypes, in a small subset of mice (HIV-WT: N=4; HIV-45: N=1; HIV-WT + 3TC: N=4; HIV-45 + 3TC: N=5) we also performed deep sequencing on 269 bp of HIV *vif* in plasma vRNA between baseline and 58 dpi. Sequencing of *vif* in HIV-WT groups revealed a minority of E45G sequences at baseline in 7/8 hu-mice (wild type E45 Vif: 98.03±2.07%; mutant E45G Vif: 1.97±2.07%) (Figure S4A-B). However, there was no significant expansion of these minority E45G sequences over time (HIV-WT 30 dpi vs 58 dpi: $p>0.9999$), nor

in response to 3TC-treatment (HIV-WT 30 dpi vs 58 dpi: p>0.9999). Conversely, at baseline, 2/6 hu-mice infected with HIV-45G had a minority of sequences that encoded wild type Vif (mouse #2256: 1/230 cDNAs, 0.4%; mouse #2322: 2/290 cDNAs, 0.7%) (Figure S4A-B) and 5/6 of these hu-mice showed expansion or emergence of wild type Vif at 58 dpi. This suggests viral adaptation over the course of infection due to selective pressure exerted by A3G in HIV-45G infected hu-mice.

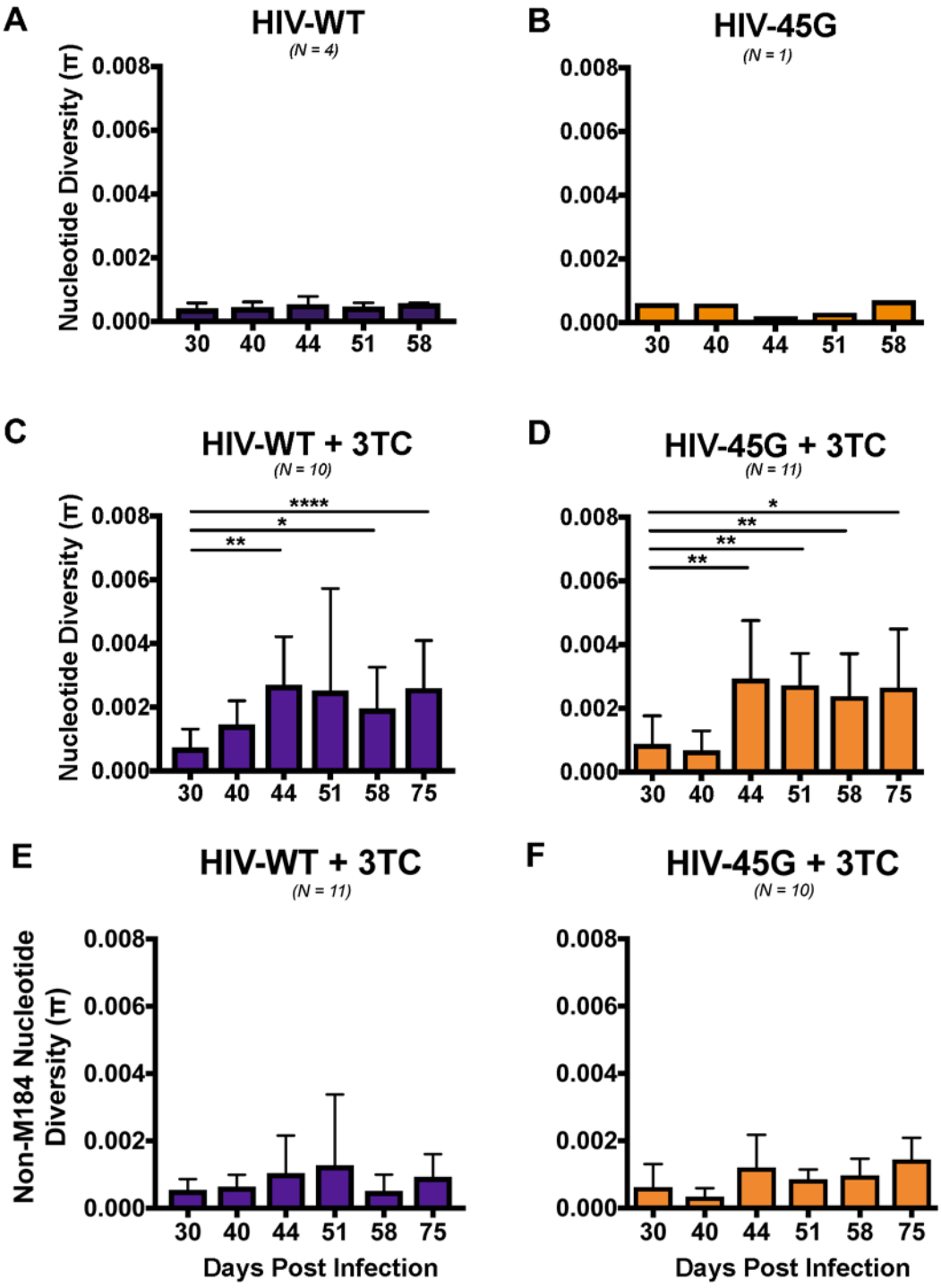


Figure 4*. 3TC treatment longitudinally drives viral diversification in plasma.* *A.-D. Nucleotide diversity over time in untreated hu-mice infected with HIV-WT (A) or HIV-45G (B) and in 3TC treated hu-mice infected with HIV-WT (C) or HIV-45G (D). E.-F. Nucleotide diversity excluding mutations at the amino acid residue 184 (M184) in 3TC treated hu-mice infected with HIV-WT (E) or HIV-45G (F).*

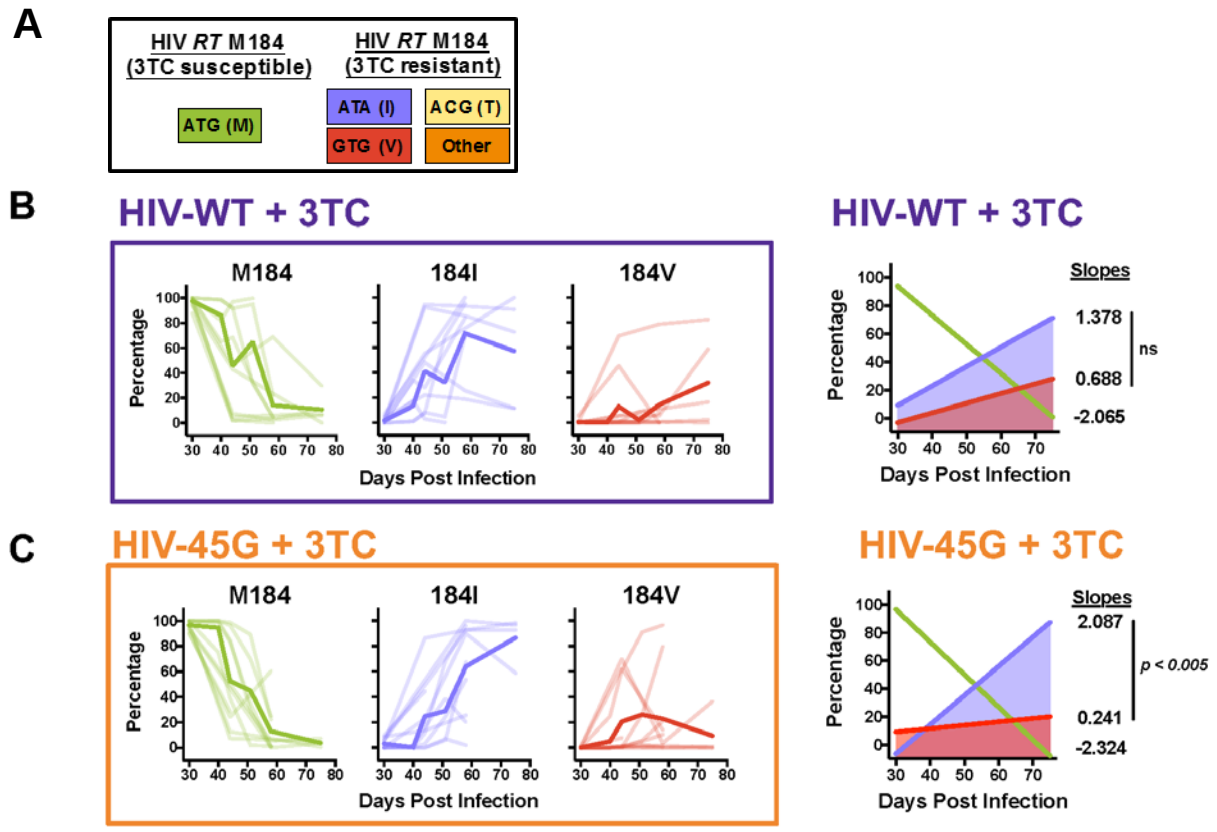


Figure 5. ***Drug resistance development in plasma of 3TC treated hu-mice.*** A. *Description and color-coding of the drug susceptible and drug resistant mutations at HIV RT codon 184.* B. *Frequencies of drug susceptible (M184) and drug resistant (184I, 184V) sequences in plasma of treated hu-mice infected with HIV-WT or HIV-45G over time.* C. *Linear regression of data shown in (B), with the best-fit values of the slopes indicated on the right. F-test for comparisons of slopes. ns = non-significant. Pale lines represent individual mice and bold lines represent group means.*

3.2.5. Diversification in the splenic viral compartments of infected hu-mice

Within a host, HIV can exhibit distinct diversification in different tissue compartments [263, 291]. In addition, a purifying selection pressure may limit the amount of diversity found in the circulating viral quasispecies compared to cell-associated proviruses [269]. Thus, we sequenced cell-associated HIV transcripts and proviruses in the spleens of hu-mice at the end of infection (i.e. 58, 75 or 85 dpi depending on the experiment, see Figure 2A). Splenocytes were available from 34 animals (HIV-WT: N=6; HIV-45G N=6; HIV-WT + 3TC: N=11, HIV-45G + 3TC: N=10). We obtained sequencing data for HIV *RT* in cell-associated HIV transcripts for all 34 spleens and sequencing data for HIV *RT* in proviruses for 21/34 spleens. This limited proviral sampling may be due to mutations that interfered with primer annealing and amplification.

In untreated HIV-WT infected hu-mice, nucleotide diversity was comparable across all nucleic acid compartments at endpoint (π, plasma vRNA: 0.00057; cell-associated transcripts: 0.00103; HIV DNA: 0.00128) (Figure 6A). In contrast, in untreated HIV-45G infected hu-mice, proviral diversity

(π=0.01347) was higher than in cell-associated transcripts (p=0.0162) and 10.5-fold greater than proviral diversity in the HIV-WT group (p=0.0377). HIV-45G proviral diversification resulted from an estimated average mutation rate of 9.06±6.54 $x10^{-3}$ substitutions/bp, in contrast to only 6.61±3.79 $x10^{-3}$ substitutions/bp in the proviruses of HIW-WT infected hu-mice. In the treated groups, proviral diversity was significantly greater than in the RNA compartments for both viruses (π plasma vRNA, HIV-WT: 0.00242, HIV-45G: 0.00274; π cell-associated transcripts, HIV-WT: 0.00247, HIV-45G: 0.00329; π HIV DNA, HIV-WT: 0.00551, HIV-45G: 0.00682), but diversity was comparable between corresponding compartments of the two virus (Figure 6B). Furthermore, as hypermutated proviral sequences (i.e. proviruses that harbor extensive G-to-A mutations in APOBEC3 target sites) are largely evolutionary dead-ends (i.e. give rise to replication incompetent proviruses) [193], they can result in an overestimation of viral diversification. Therefore, we used the online tool Hypermut 2.0 to identify hypermutant proviruses (see methods section) and thereby quantify proviral nucleotide diversity excluding hypermutants. In untreated mice, hypermutants contributed to 24.4±21.2% and 52.2±30.4% of proviral diversity in HIV-WT and HIV-45G mice (p=0.3429), respectively. In treated mice, hypermutants contributed to 27.9±19.2% of HIV-WT proviral diversity and 24.2±28.6% of HIV-45G proviral diversity (p=0.6964).

To determine what is driving the viral diversification, we measured the fraction of GG dinucleotides mutated to AG and GA dinucleotides mutated to AA in the different quasispecies compartments (Figure 6C-F). These analyses revealed that in both untreated and treated mice, diversification was driven by GG-to-AG mutations, with only few GA-to-AA mutations observed. In untreated hu-mice, HIV-45G proviruses carried a greater fraction of GG-to-AG mutations (13.3%) than HIV-WT mice (0.5%, p=0.0401, Figure 6C). Strikingly, in the context of 3TC, this difference was lost (HIV-45G: 9.3% vs HIV-WT: 3.9%, p=0.1998, Figure 6D). As G-to-A mutations, particularly those in GG-to-AG or GA-to-AA dinucleotide contexts may result in premature stop codons, we also quantified the number of putative stop codons per 1000 codons sequenced (Figure 6G-H). Overall, the number of stop codons were highest in the proviral compartments compared to the RNA viral compartments, in line with a purifying selection of hypermutated proviruses. In untreated hu-mice, there was less than one stop codon per 1000 sequenced codons in all three HIV-WT compartments (plasma vRNA: 0.01; cell-associated transcripts: 0.05; proviruses: 0.17), whereas the HIV-45G proviruses and cell-associated transcripts contained a significantly greater number of putative stop codons than in the HIV-WT counterparts (proviruses: 8.3, p=0.0442; cell-associated transcripts: 0.92, p=0.0022). In treated hu-mice, stop codon rates in proviruses were comparable (p=0.3660) in HIV-WT (1.1 per 1000) and HIV-45G (6.1 per 1000) groups, but the number of stop codons was greater in HIV-45G cell-associated transcripts than those of HIV-WT (HIV-WT: 0.1, HIV-45G: 0.6; p<0.0001) (Figure 6H). This discrepancy may be due to a limited sampling of the proviral quasispecies. Taken together, these results suggest A3G-driven mutagenesis contributes strongly to viral diversification, but that increased mutagenesis can also results in higher levels of

defective proviruses and HIV transcripts. Footprints of A3G mutagenesis were prominent at the proviral level of untreated HIV-45G hu-mice, whereas the selective pressure (i.e. bottleneck) imposed by 3TC may be masking differences in the treated groups.

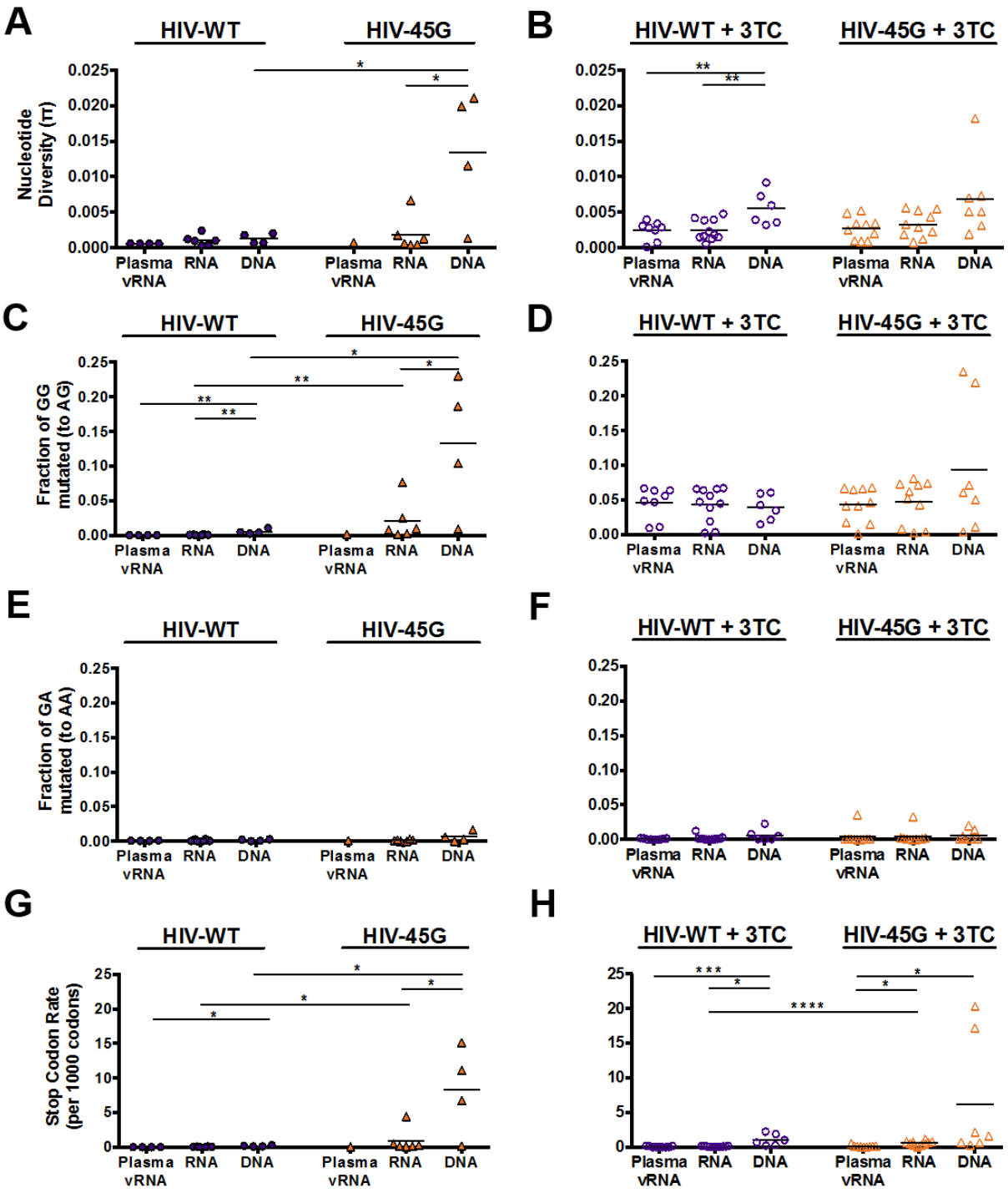


Figure 6. ***Characterization of quasispecies in the plasma and spleen compartments of hu-mice.*** A.-B. *Nucleotide diversity of quasispecies found in the plasma and spleen compartments in untreated (A) and 3TC treated (B) hu-mice at the end of infection.* C.-F. *Fractions of GG dinucleotides mutated to AG, or GA dinucleotides mutated to AA in plasma and spleen compartments in untreated (C, E) and treated mice (D, F).* G.-H. *Numbers of codons mutated to stop codons for every 1000 codons sequenced in viral compartments in untreated (G) and treated hu-mice (H).*

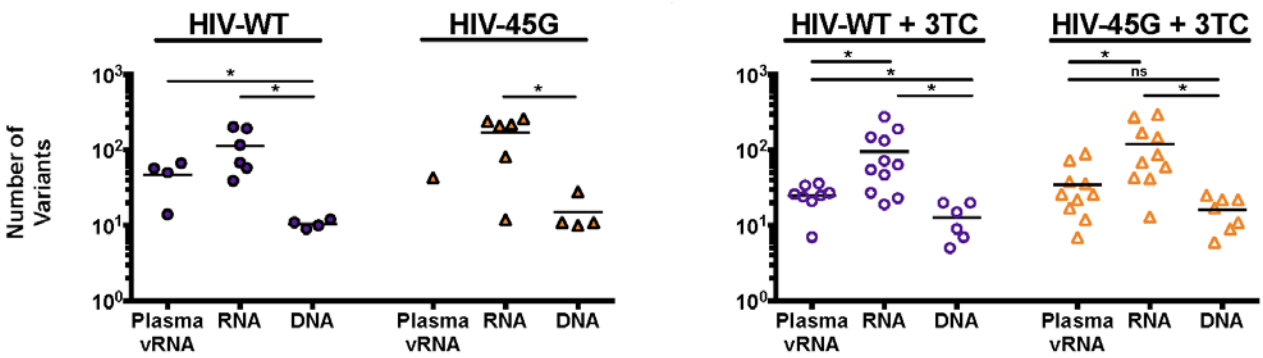


Figure 7. Number of unique variants across the plasma and spleen viral compartments. *Numbers of distinct variants in untreated (I) and (J) treated hu-mice at the end of infection.*

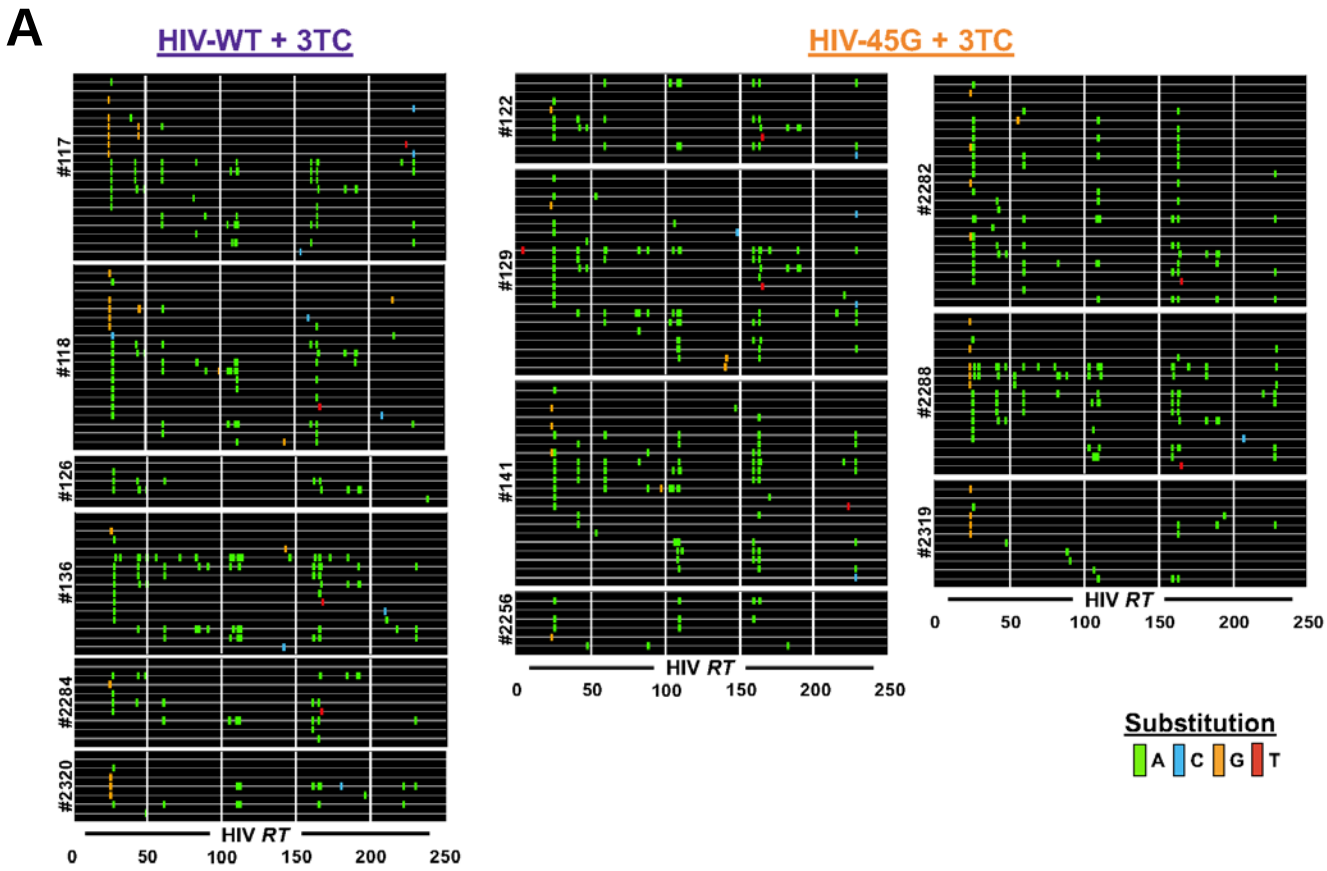


Figure 8. ***Characterization of proviral variants in treated hu-mice.*** A. *Depiction of the mutational landscape of proviral variants in treated hu-mice. The individual horizontal lines represent distinct proviral sequences and substitutions are depicted as colored bars at affected positions.*

B

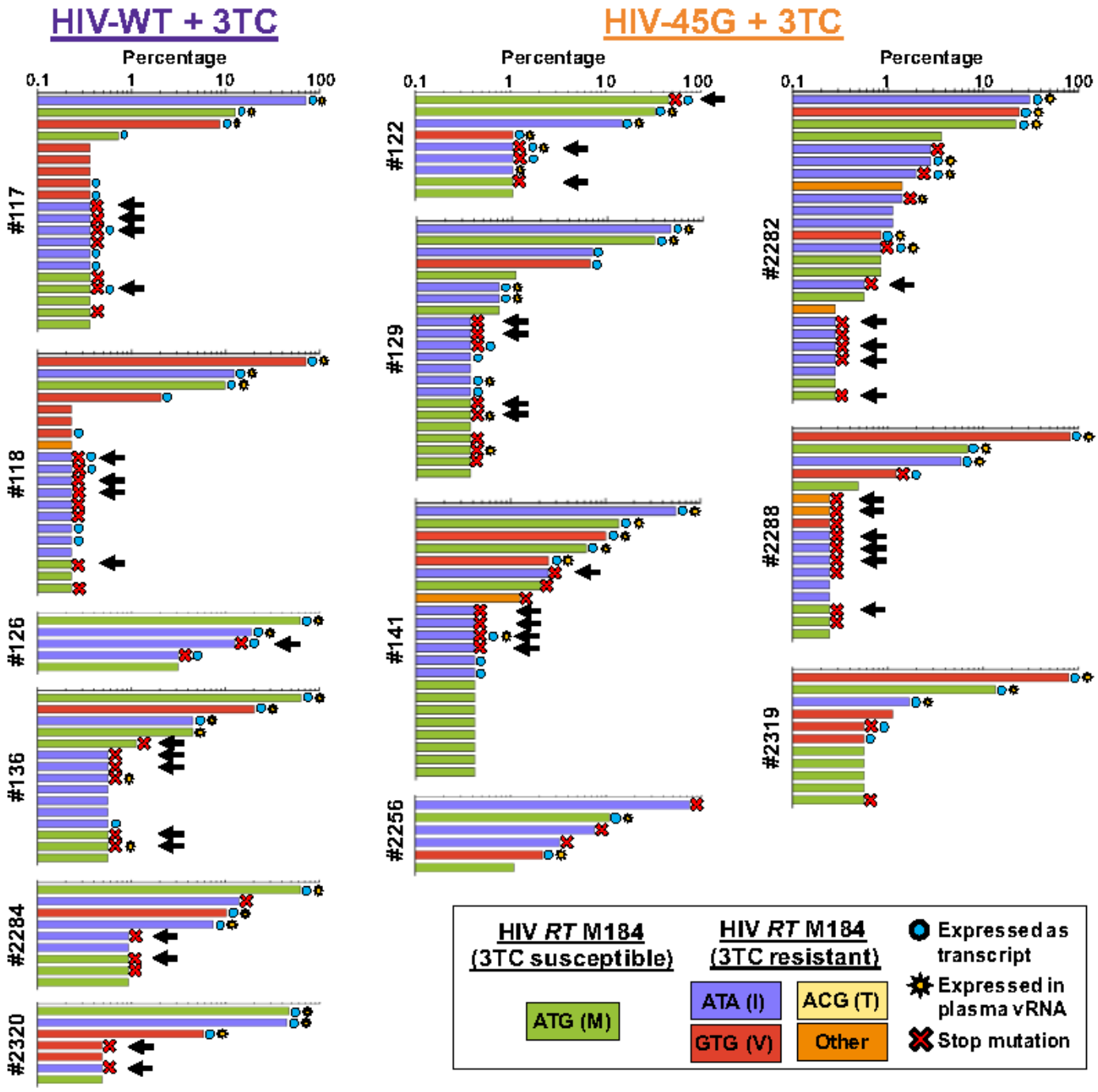


Figure 8. (Cont.) ***Characterization of proviral variants in treated hu-mice.*** B. *Frequencies of proviral variants in treated hu-mice. Drug susceptible or resistance mutations are color-coded. Proviral variants represented in splenocyte transcripts (blue circle) or plasma vRNA (yellow star) are indicated. Variants with premature stop mutations are indicated (red X). Hypermutants (see methods) are indicated by black arrows.*

3.2.6. Characterization of viral landscapes found in the plasma and splenic viral compartments of infected hu-mice

To understand the dynamics between archival (DNA) and expressed (RNA) viral landscapes in the spleen and plasma of infected hu-mice, for each animal we identified, quantified and characterized the viral variants present in each compartment (Figures 7-9). Herein, a viral variant corresponds to a distinct viral population characterized by *RT* sequences that are identical in sequence but have a different UMID. Data for all three viral compartments were available for 17/21 hu-mice (for the remaining four hu-mice only cell-associated HIV RNA and DNA data were available). We identified a total of 2948 distinct HIV *RT* variants. The number of variants differed starkly between viral nucleic acid compartments (Figure 7). Strikingly, the highest number of variants was observed in cell-associated HIV transcripts (HIV-WT: 113±71/sample; HIV-45G: 171±100; HIV-WT + 3TC: 96±82; HIV-45G + 3TC: 120±100) and overall proviruses contained the lowest number of variants, displaying about one order of magnitude fewer variants than cell-associated transcripts (untreated: HIV-WT: 11±2, HIV-45G: 15±9; treated: HIV-WT: 13±7, HIV-45G: 16±8). These results are unexpected and we lack a plausible biological explanation to explain these. Given that in acutely HIV-infected patients, PBMCs were shown to principally harbor defective proviruses [47], we expect a greater number of variants in the proviruses, which are then absent in the transcribed viral populations due to purifying selection of defective viruses. Thus, we postulate that these unexpected results may be the consequence of limited recovery of splenic proviral quasispecies, especially as we observed a greater divergence between proviral variants than between the RNA variants (i.e. proviral variants generally differed by several mutations, whereas the majority of RNA variants differed by only one base pair).

We observed distinct mutations reflecting extensive adenosine substitutions in a dinucleotide context suggestive of APOBEC3 activity in the proviral variants of treated hu-mice (Figure 8A). Interestingly, despite antiretroviral pressure, we found both drug susceptible and drug resistant proviruses in treated hu-mice (Figure 8B). The 3TC susceptible variants were often intact without premature stop codons (HIV-WT: 15/25 M184 variants; HIV-45G: 34/45 M184 variants). These drug-susceptible proviruses probably reflect the quasispecies present prior to 3TC treatment. Overall, in the proviruses of treated hu-mice, there were comparable levels of drug susceptible sequences (HIV-WT: 46.19±27.03%; HIV-45G: 28.06±25.28%, p=0.2373) as well as drug resistant M184I (HIV-WT: 33.31±24.03%; HIV-45G: 40.13±32.95%, p=0.6831) and M184V (HIV-WT: 20.46±27.41%; HIV-45G: 31.31±38.93%, p=0.5796) sequences. In addition, the numbers of proviruses with premature stop codons were comparable between both groups (HIV-WT: 37.2±5.8%; HIV-45G: 42.5±13.0%; p=0.3786). Furthermore, to identify potential evolutionary bottlenecks (i.e. selective pressure imposed by 3TC and purifying selection), we mapped the fraction of proviruses that gave rise to cell-associated transcripts and plasma virions in the treated

hu-mouse groups (Figure 8B). In the absence of bottlenecks, every provirus would be transcribed with equal efficiency, resulting in equal numbers of cell-free virions. However, we observed that only 46.3% (HIV-WT) and 39.8% (HIV-45G) of proviruses were transcribed (p=0.4924), with 31% (HIV-WT) and 32% (HIV-45G) being represented in cell-free virions (p=0.8742). We observed fewer proviral variants expressed in plasma compared to the number expressed as cell-associated transcripts only for HIV-45G infected hu-mice (p=0.0387). This indicates that HIV-45G proviruses were less likely to produce circulating viruses, possibly due to APOBEC3-driven mutagenesis resulting in lethal defects at the protein level.

Furthermore, to determine whether variants were comparably distributed among the cell-free and cell-associated compartments, for each hu-mouse, correlation analyses were performed for all variants found in that animal, in each compartment (Table 2). If the correlation coefficient was greater than 0.6 (R^2>0.60), we defined the correlation between the two nucleic acid compartments as strong. There was a strong correlation between the distribution of variants in proviruses and cell-associated transcripts in 14/21 hu-mice (HIV-WT: 4/4; HIV-45: 2/4; HIV-WT + 3TC: 3/6; HIV-45G + 3TC: 5/7) (Table 2 and Figure 9). The same hu-mice showed a strong correlation between proviral and plasma vRNA variant distributions. Additionally, all distributions within cell-associated transcripts correlated very strongly (i.e. mean R^2=0.96±0.07) with plasma vRNA. These findings are suggestive of the existence of a purifying selection from proviruses to plasma virions, but also that there is good mixing between the splenic and plasma viral compartments studied.

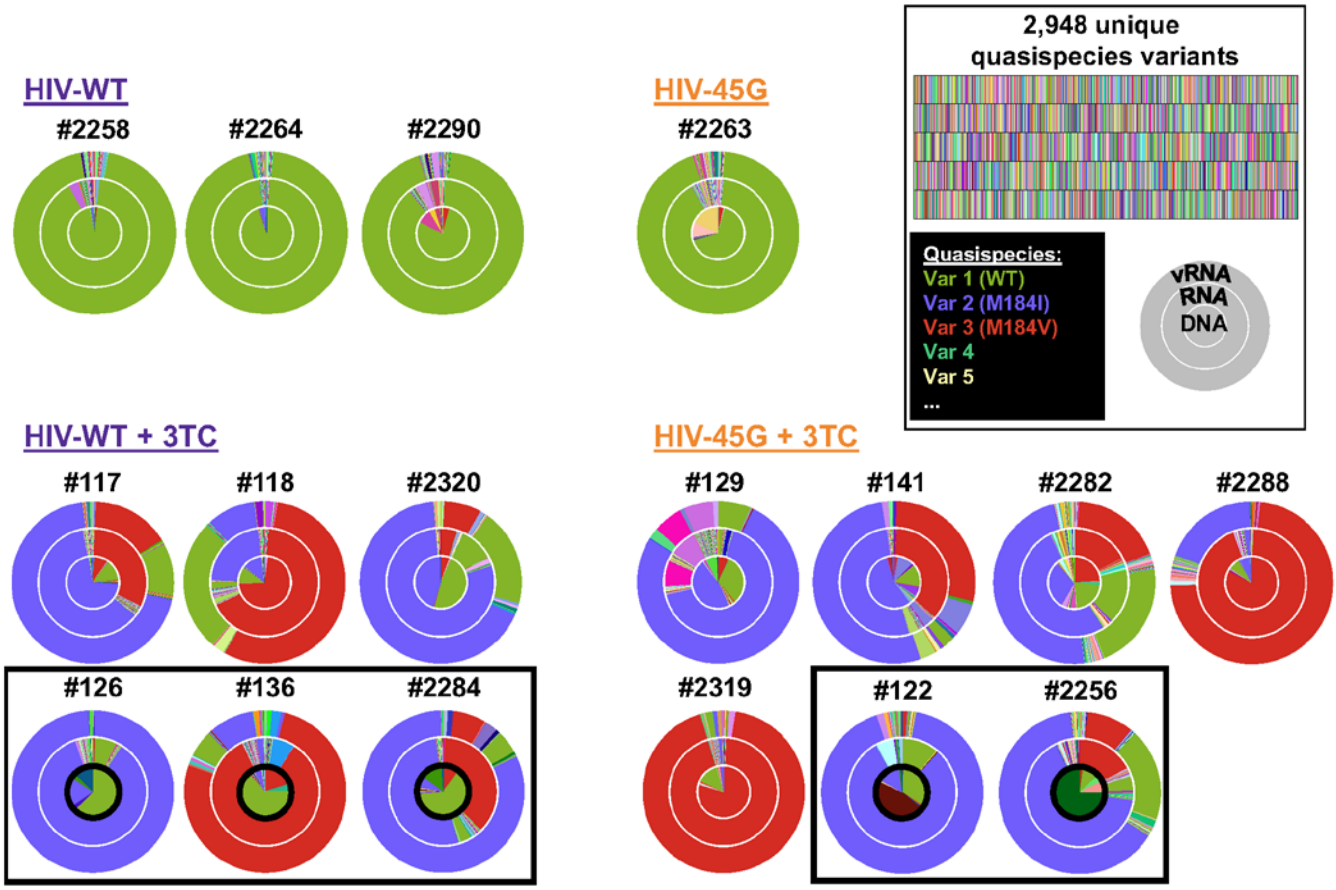


Figure 9. ***Viral quasispecies landscapes in splenocyte and plasma compartments.*** *From inside to outside, concentric circles depict the frequencies of distinct quasispecies variants among splenocyte-associated proviruses (DNA), splenocyte-associated HIV transcript (RNA) and plasma viral RNA (vRNA) landscapes. Each of 2948 unique quasispecies sequences is indicated by a distinct color and the 5 most frequent variants (and nucleotide mutations) are labeled by their respective colors (see inset). Proviral landscapes that weakly correlate with viral RNA landscape are outlined in black.*

		DNA v. RNA			DNA v. Plasma vRNA			RNA v. Plasma vRNA		
	Mouse	Slope (mean±SD)	R^2	p-val	Slope (mean±SD)	R^2	p-val	Slope (mean±SD)	R^2	p-val
HIV-WT	2258	0.9542 ± 0.0008229	0.9978	<0.0001	0.9942 ± 0.0005749	0.999	<0.0001	1.041 ± 0.0007026	0.999	<0.0001
	2264	1.024 ± 0.0007567	0.9984	<0.0001	1.005 ± 0.0007697	0.9983	<0.0001	0.9813 ± 0.0001899	1	<0.0001
	2290	1.106 ± 0.002783	0.9817	<0.0001	1.17 ± 0.003008	0.9809	<0.0001	1.057 ± 0.0006288	0.999	<0.0001
	8	1.061 ± 0.00141	0.9948	<0.0001	N/A	N/A	N/A	N/A	N/A	N/A
HIV-45G	2263	1.242 ± 0.007055	0.9131	<0.0001	1.292 ± 0.00735	0.9129	<0.0001	1.04 ± 0.0003693	1	<0.0001
	4	0.8174 ± 0.0187	0.3931	<0.0001	N/A	N/A	N/A	N/A	N/A	N/A
	9	1.47 ± 0.03704	0.348	<0.0001	N/A	N/A	N/A	N/A	N/A	N/A
	32	1.09 ± 0.001689	0.993	<0.0001	N/A	N/A	N/A	N/A	N/A	N/A
HIV-WT + 3TC	117	0.8677 ± 0.006641	0.8527	<0.0001	0.9721 ± 0.001906	0.9888	<0.0001	0.9959 ± 0.00554	0.916	<0.0001
	118	0.9392 ± 0.00329	0.9651	<0.0001	0.8057 ± 0.004713	0.9084	<0.0001	0.8016 ± 0.006872	0.822	<0.0001
	126	0.4526 ± 0.02224	0.1232	<0.0001	0.4188 ± 0.02622	0.07962	<0.0001	1.148 ± 0.001607	0.994	<0.0001
	136	0.379 ± 0.02188	0.09234	<0.0001	0.4257 ± 0.01957	0.1383	<0.0001	0.9104 ± 0.002144	0.984	<0.0001
	2284	0.2199 ± 0.01739	0.05142	<0.0001	0.2185 ± 0.0224	0.03126	<0.0001	1.093 ± 0.01207	0.735	<0.0001
	2320	0.9396 ± 0.01404	0.6029	<0.0001	0.9242 ± 0.009942	0.7456	<0.0001	0.8732 ± 0.002593	0.975	<0.0001
HIV-45G + 3TC	122	0.4587 ± 0.02415	0.109	<0.0001	0.406 ± 0.02809	0.06613	<0.0001	1.124 ± 0.00308	0.978	<0.0001
	129	0.9688 ± 0.01246	0.6722	<0.0001	1.169 ± 0.01295	0.7343	<0.0001	1.149 ± 0.002062	0.991	<0.0001
	141	0.8802 ± 0.0116	0.6614	<0.0001	0.9041 ± 0.009143	0.7683	<0.0001	0.9448 ± 0.002301	0.983	<0.0001
	2256	0.01344 ± 0.01638	0.0002283	0.4119	0.03723 ± 0.01627	0.001772	0.0222	0.9646 ± 0.004446	0.941	<0.0001
	2282	1.153 ± 0.009142	0.8436	<0.0001	1.161 ± 0.007129	0.9	<0.0001	0.9677 ± 0.002193	0.985	<0.0001
	2288	1.125 ± 0.002077	0.99	<0.0001	0.8866 ± 0.003344	0.9597	<0.0001	0.7781 ± 0.003476	0.944	<0.0001
	2319	1.152 ± 0.00363	0.9716	<0.0001	1.135 ± 0.003307	0.9756	<0.0001	0.9827 ± 0.0004018	1	<0.0001

Table 2. ***Correlations between splenocyte-associated proviruses and HIV transcripts, and plasma viral RNA landscapes.***

3.3. Discussion

Evidence of APOBEC3 G-to-A hypermutation footprints in the HIV genomes and Vif variants with suboptimal APOBEC3-neutralization activities in patients [187-189] suggest that the APOBEC3G-Vif axis contributes to viral diversification, yet this theory remains highly debated. In this study, we used humanized mice infected with isogenic HIV clones differing in their ability to counteract A3G, but displaying comparable replication in the absence of A3G *in vitro*, to dissect the effects of A3G-mediated mutagenesis on viral replication, diversification and adaptation to the nucleoside analog reverse transcriptase inhibitor lamivudine (3TC). We first characterized the phenotype of the viruses either in natural (i.e. untreated) infections or in the context of 3TC monotherapy by longitudinally quantifying viremia and PBMCs subsets. To gain insights into the mechanisms driving viral fitness and evolution, we sequenced the viral reverse transcriptase gene (HIV *RT*) from cell-free plasma virions, as well as cell-associated viral transcripts and proviruses in the spleens of infected hu-mice, allowing us to characterize quantitatively the viral quasispecies present longitudinally in the plasma and at endpoint in the splenocytes.

In this study, we observed clear differences in fitness between the Vif-mutant HIV with only partial (20%) activity against A3G (HIV-45G) and the wild type Vif HIV (HIV-WT; 95% activity against A3G) replicating in hu-mice, and characterization of the viral quasispecies in the plasma and splenic compartments provided evidence of A3G's underlying role in these phenotypic manifestations. Specifically, we found that HIV-45G established a robust infection in hu-mice, but over time exhibited attenuated viral replication compared to HIV-WT (Figure 3A-B). This attenuated fitness phenotype was characterized by a 10.5-fold greater diversity in the splenic proviruses of hu-mice infected with HIV-45G compared to those infected with HIV-WT (Figure 6A). Importantly, the proviral diversity in HIV-45G infected hu-mice was predominantly sustained by GG-to-AG mutations (i.e. characteristic A3G mutational footprint), which resulted in a ~48-fold greater stop codon rate (i.e. number of stop codons per 1000 codons) than that observed in the proviral compartment of HIV-WT infected hu-mice (Figure 6C, G). Hypermutated proviruses contributed to approximately half of the proviral diversity seen in the HIV-45G hu-mice. Interestingly, some hypermutated proviral variants were also expressed as splenocyte-associated transcripts and circulating virions. These results suggest that partially neutralized A3G in HIV-45G infections contributes to a more diversified pool of infectious viruses, in line with other studies [192, 267]. Moreover, the proviral mutation rate observed in HIV-45G infected mice (~$9x10^{-3}$ substitutions/bp) is much higher than the estimated *in vitro* error rate of HIV RT ($3 x10^{-5}$) [20, 174, 175], making it difficult to attribute this high level of GG-to-AG substitutions to RT-error alone. In addition, the high HIV-45G proviral mutation rate observed in this study is comparable to mutation rates found in treatment-naïve patient PBMCs ($4.1 \pm 1.7 x10^{-3}$), where both RT- and APOBEC3-associated mutagenesis were identified [292]. Finally, characterization of the distribution of viral variants between proviruses (HIV DNA) and cell-associated viral transcripts (HIV RNA) in the spleen, and

cell-free virus in the plasma (plasma vRNA), revealed a tendency for HIV-45G proviruses to be underrepresented in the transcribed viral pools (Table 2 and Figures 8-9). These findings are in line with a purifying selection of hypermutated genomes occurring from proviral DNA, to cell-associated viral RNA and finally cell-free viral RNA, such that only viable viruses mostly devoid of hypermutation are released into the plasma [166, 269, 292, 293]. Indeed, in the plasma quasispecies we observed comparable nucleotide diversity and mutation rates for both viruses over time, with no evidence for the accumulation of mutations or increase in diversity over time, despite extensive footprints of G-to-A mutations in splenic proviruses in HIV-45G infected hu-mice (Figures 4, 7A and S3). Previously, our collaborators showed that hypermutated proviruses can be rescued by recombination (i.e. template switching during reserve transcription) between heterogonous RNA genomes in virions produced in co-infected cells *in vitro* [179]. Thus, recombination between hypermutated and intact genomes may be a mechanism by which hypermutated proviruses contribute to viral diversification *in vivo*. Furthermore, the transcription and translation of portions of defective proviruses has been documented in patients [294] and such proviruses may contribute to HIV pathogenesis by chronically stimulating host-immune responses [295, 296]. Taken together these results indicate that A3G-mediated mutagenesis diversifies the proviral repertoire, with diverging consequences for HIV. As proviruses with A3G-mutagenic footprints are often defective (hypermutated) and cannot be properly transcribed or packaged into virions, A3G-mediated mutagenesis attenuates viral fitness over time. Yet these hypermutated genomes may still contribute to viral evolution through recombination, as well as sustain an immune environment favorable for HIV replication (i.e. activation of resting CD4+ T cells).

To assess the contribution of A3G-mediated mutagenesis to HIV adaption, we introduced a selective pressure into our model by treating infected hu-mice with 3TC, one month post-infection. As 3TC-resistance is solely conferred by substitution of the methionine codon 184 (M184) in HIV *RT* predominantly with an isoleucine (M184I) or a valine (M184V), and the M184I variant results from a G-to-A mutation in A3G's preferred dinucleotide context (i.e. GG-to-AG), this anti-HIV drug is well suited to study whether A3G plays a role in HIV evolution *in vivo.* Specifically, if A3G significantly contributes to viral adaptation, we expect to see differences in the emergence of the M184I variant between hu-mice infected with HIV-WT and HIV-45G. Indeed, we found that in hu-mice, although both viruses were initially comparably sensitive to restriction by 3TC, the partially defective Vif mutant HIV adapted faster to 3TC and was thereafter fitter than HIV-WT (Figure 3C-F). This suggests that *in vivo*, suboptimal neutralization of A3G confers an enhanced evolutionary potential onto the virus. Analysis of the plasma quasispecies revealed that in both HIV-WT and HIV-45G infected hu-mice, there were comparably low levels of drug resistant variants at baseline (i.e. prior to the start of 3TC therapy) and that rebound in viremia was concurrent with the emergence of the drug-resistant M184I/V variants in the plasma, with a predominance of the M184I mutation (Figure 5). These findings mirror what is seen in patients on 3TC monotherapy

[277, 278]. Importantly, in the hu-mice infected with HIV-45G, we observed an 8.3-times faster emergence of the M184I variant compared to the M184V variant. This was not the case for hu-mice infected with HIV-WT. As the M184I variant is prone to arise from A3G-mediated mutagenesis, this further supports the notion that A3G can boost viral adaptation *in vivo*. Although a study by Keulen et al. attributed the early appearance of the M184I variant to HIV RT's G-to-A mutational bias [273], in this study A3G's GG-to-AG mutagenic effect on HIV was not considered. Also, since then, our collaborators have shown that A3G specifically drives the emergence of the M184I mutation (and not the M184V or other minority 3TC-resistant variants) in proviruses in *ex vivo*-infected PBMCs [179], whilst other groups have shown that HIV's RT has a bias for GA targets [20, 272, 274].

Theoretically, we expect higher viral diversity in hu-mice infected with HIV-45G and treated with 3TC since HIV's high rate of error prone replication, as well as A3G's mutagenic activity would be contributing to it. Surprisingly, diversification in the cell-free plasma vRNA was comparable between both viruses, increasing over the first two weeks of treatment and then remaining stable until the end (Figure 4). Plasma vRNA diversity was largely a result of GG-to-AG mutations (Figure S3). Two mechanisms may be affecting diversification in *RT* in the circulating viral pool and masking differences between the viruses: the purifying selection of hypermutated genomes and the bottleneck imposed specifically on RT by 3TC. The overall increase in diversity seen in the treated groups compared to the untreated hu-mice groups supports this notion, as well as the fact that diversity was largely (~40%) driven by mutagenesis at residue 184 in HIV RT (Figure 4). Furthermore, we observed an emergence and expansion of WT *vif* in the plasma quasispecies of HIV-45G infected hu-mice (Figure S4), suggesting a selective pressure from A3G.

Moreover, our genotypic analyses first focused on plasma quasispecies as these can readily be sampled longitudinally, however, the main sites of HIV replication occur in lymphoid tissues [51], such as the spleen, which we can only access when we sacrifice the mice at the end of the experiments (i.e. endpoint). Thus, we also sequenced the cell-associated proviral and HIV transcript quasispecies present in splenocytes and compared these to the cell-free plasma viral quasispecies at endpoint. Surprisingly, nucleotide diversity, mutation rates and number of variants were comparable between the two viruses across all corresponding compartments at endpoint (Figures 6-7). Both HIV-WT and HIV-45G proviruses were predominantly and comparably, affected by GG-to-AG mutations. In addition, comparable proportions of hypermutated sequences contributed to proviral diversification in both groups and unlike in the untreated groups, premature stop codons were not more common in the HIV-45G proviruses than the HIV-WT proviruses. Thus, it is possible that in the context of 3TC treatment, A3G's contribution to diversification is masked by the bottleneck imposed on the *RT* gene by 3TC [297]. Indeed, Mansky et al showed *in vitro* that 3TC treatment increases HIV's mutation frequency [298]. Alternatively, as we showed that for both

viruses, less than 47% of proviruses were transcribed and less than 33% were found in the circulating viral pool (Figures 8-9), it is likely that a purifying selection of hypermutated genomes limits our ability to capture the extent of A3G-induced diversification when we can only longitudinally characterize cell-free plasma quasispecies. For example, Cuevas et al, showed that in treatment-naïve patients, mutation rates in cell-free plasma quasispecies were 44-times lower than in the PBMC-associated proviruses, with a large proportion of hypermutated viral genomes not reaching the plasma [292]. As we could only assess cell-associated viral RNA and DNA quasispecies in the spleen at one time point (well after 3TC-resistant variants predominated), we cannot exclude the possibility that, like in plasma, in the context of 3TC, HIV-45G only has superior kinetics of viral diversification in the spleen. In this regard, we speculate that a longitudinal characterization of cell-associated quasispecies in PBMCs and cross-sectional study of spleen quasispecies may provide more information on how A3G shapes viral diversification. However, given the limited volume of blood that can be collected in hu-mice, in this study it was not possible to include such an analysis of PBMCs. It may be that even with a modified sequencing workflow, the amount of PBMCs that can be serially collected from hu-mice would not suffice for accurate and high-resolution sequencing of the viral quasispecies therein.

Furthermore, as viruses in diverse anatomical compartments have been shown to influence one-another [291, 299, 300], we cannot rule out the impact of proviral landscapes seeded in other tissues and cellular reservoirs, on the circulating plasma quasispecies and the phenotypes observed. For example, in secondary lymphoid tissues, such as lymph nodes and gut-associated lymphoid tissue, replication dynamics and modes of transmission influence distinct proviral landscapes in patients and hu-mice [92, 299, 301, 302]. Indeed, quantitative comparisons of viral variants in the plasma and splenocytes of 3TC-treated hu-mice at least partly support the notion that viral replication in other tissues, such as in lymph nodes, gut or bone marrow, may be shaping viral evolution in our model. Specifically, although the viral landscapes in the different compartments correlated strongly in the majority of animals (Table 2), we found that splenic proviral landscapes can be distinct from the cell-free plasma compartments (Figure 9). Thus, viral quasispecies from other tissue compartments could be influencing the plasma viral repertoire and thereby explain the different adaption phenotypes observed for the two viruses.

Lastly, APOBEC3-induced mutations often result in defective proviruses (i.e. contain premature stop codons or inactivating hypermutation) incapable of encoding for a replication-competent virus [47, 114]. Yet, surprisingly, rather than being evolutionary dead-ends, we show that some hypermutated proviruses are not only transcribed in infected splenocytes, but can be packaged and found in circulating virus (Figure 8B). The role of hypermutated quasispecies *in vivo* has yet to be fully elucidated, but *in vitro* studies have found virions containing aberrant stop codons within gag [269] and shown that defective viral sequences can contribute to sequence diversity and

resistance mutations through recombination events during coinfection [179]. Therefore, we postulate that a more diverse proviral landscape resulting from incomplete neutralization of A3G, even if littered by hypermutated proviruses, can increase the potential number of viral variants that can contribute to adaptation. Additionally, as discussed earlier, defective proviruses may also drive chronic immune activation, which is favorable to HIV replication [296].

In conclusion, we demonstrate that suboptimal neutralization of A3G by a partially defective Vif variant improves the fitness of HIV in the presence of a selective pressure. Using high-resolution sequencing of plasma and splenocyte compartments, we show that this phenotype results from increased diversification in proviral and circulating quasispecies, characterized by an A3G-mutational footprint. Moreover, distinct dynamics and quasispecies landscapes between plasma, spleen and other viral compartments may influence systemic viral fitness and require further investigation. We believe the depth and accuracy of our sequencing approach are key and distinguish our work from previous studies investigating the role of A3G in humanized mice [166, 263].

3.4. Materials and Methods

Ethics statement: Animal experiments were approved by the Cantonal Veterinary Office (#26/2011 & #93/2014) and performed in accordance to local guidelines (TschV, Zurich) and the Swiss animal protection law (TschG). Procurement of human cord blood was approved by the Ethical Committee of the University of Zurich (EK #EK1103) and cord blood was collected with written informed consent from donors.

Cells and reagents: Cell lines were obtained from NIH Aids reagent program. 293T cells and TZM-bl cells were cultured in DMEM medium (Lonza) supplemented with 10% FBS (Lonza), 1% Penicillin-Streptomycin (Life technologies). MT4 cells and primary mononuclear cells (MNCs) derived from cord blood were cultured in RPMI medium (Lonza) supplemented as above. MNCs were activated with PHA at 1ug/mL for 48 hours prior to use in infection assays. 3TC tablets (Epivir, GlaxoSmithKlein, UK) were weighed, pulverized and admixed to food pellets as previously described [163].

Viral stocks: Viral stocks were produced by transient transfection of 293T cells with polyethylenimin (Sigma-Aldrich) and the pNL4-3 plasmid (NIH Aids reagent program, Cat. No. 114), or plasmids encoding the various Vif-mutants (HIV-45G and HIV-ΔSLQ) within the backbone of pNL4-3. The HIV-45G and HIV-ΔSLQ mutants were generated as previously described [186, 188, 303]. After 3 to 18 hours incubation, the supernatant was exchanged for fresh D-10 medium. Forty-eight hours post-transfection, the supernatant was harvested, filtered (0.22um, Steriflip, EMD Millipore), aliquoted and frozen at -80°C until use. The tissue culture infectious dose 50 (TCID50) of the viral stocks were determined by infection of TZM-bl cells seeded in 96-well plates (10^4 cells/well) with serial dilutions of the viral stocks. Forty-eight hours post-infection the cells were washed, lysed and fixed and β-galactosidase measured in the plates using the Galacto-Star B-Galactosidase Reporter Gene System (Applied Biosystems). Plates were scored for the number of negative wells and TCID50 was then calculated using the Spearman-Karber formula.

Generation of humanized mice: Hu-mice were generated as previously described [163]. Briefly, newborn immunodeficient NOD-scid IL-2Rγ-null (NSG) mice (Jackson laboratory) were sub-lethally irradiated (1 Gy) and transplanted intrahepatically with approximately $2.0\pm0.5\times10^5$ cord blood-derived CD34+ cells. Twelve to sixteen weeks after transplantation, human engraftment and *de novo* human immune system reconstitution in the mice were assessed by flow cytometry (CyAn™ ADP Analyzer, Beckman Coulter) by staining PBMCs with anti-human monoclonal antibodies against the following cell-surface markers: CD45-Krome Orange (Beckman Coulter), CD3-PE, CD4-Pe Cy7, CD8-BV421, CD19-APC (all from Biolegend). Animals were housed under specific pathogen free conditions.

Infection and 3TC treatment of humanized mice: Hu-mice were infected intraperitoneally with the various HIVs at a dose of 2×10^5 TCID50 per mouse. Plasma viremia was monitored 30 days post-infection and thereafter on a biweekly basis using the Cobas® Amplicor technology (Roche, Switzerland) (Figure S1A). The detection limit of the assay is 400 HIV RNA copies/mL. At 30 days post-infection (i.e. baseline), some infected hu-mice were treated with 3TC. Only hu-mice with a detectable viremia at baseline were included in the analyses. At the end of the experiments, hu-mice were euthanized and spleens were harvested, processed and stored frozen in 10% DMSO/FBS in liquid nitrogen.

Amplification of HIV from plasma viral RNA: Viral RNA was extracted from 200µL frozen plasma samples using the QIAamp Viral RNA Minikit (Hilden) as per manufacturer's instruction. For amplification of samples for deep sequencing, cDNAs were synthesized using custom primers in a reverse transcription reaction. (Refer to Table S1 for primer sequences; Integrated DNA Technologies). From 5' to 3', our first generation *RT* cDNA primer (4372) included a 16 bp HIV-specific sequence (accession: AF324493.2: 3336-3351), 8 bp randomized sequence (UMID) and 23bp HIV-specific sequence (AF324493.2: 3305-3327). Viral cDNA was synthesized using the ThermoScript RT-PCR System (Thermo Fisher Scientific). Briefly, RNA and primer were denatured at 65°C for 5 min. Thermoscript reaction mix was added to the RNA and primer (50°C for 60 min, 85°C for 5 min). Viral cDNA was column purified (Zymo Clean and Concentrator) and amplified in a first round PCR using primers 1922 (AF324493.2: 2929-2946) and 1923 (AF324493.2: 3337-3356). First round PCR used Pfx50 polymerase (94°C for 2 min, followed by 23 cycles of 93°C for 15 sec, 48°C for 30 sec, 68°C for 60 sec and a final extension of 68°C for 10 min) (Thermo Fisher Scientific). First round products were purified (Zymo Clean and Concentrator) and a second PCR was performed to add Illumina-based adapters using custom primers 1690 and one of 734 primers with distinct MiSeq barcode identifiers (Table S1). The second round PCR used PfuUltra II Fusion HS Polymerase (95°C for 2 min, followed by 25 cycles of 93°C for 20 sec, 50°C for 20 sec, 72°C for 15 sec and a final extension of 72°C for 10 min) (Agilent Technologies). Second round products were confirmed by electrophoresis and purified by SPRIselect magnetic bead selection (Beckman Coulter).

To amplify *vif* sequences from plasma vRNA, a distinct cDNA synthesis primer was used (4373). From 5' to 3', primer 4373 included a 16 bp T7 sequence, an 8bp randomized sequence and 23 bp HIV-specific sequence (AF324493.2: 5386-5406). Viral cDNA was synthesized using the ThermoScript RT-PCR System and purified as described above. cDNAs were amplified in the first round PCR using primers 2402 (AF324493.2: 4993-5012) and 2401 (T7). First round products were purified and amplified in a second round PCR using primers 2403 and one of 80 primers with

distinct MiSeq barcode identifiers and complementarity to the T7 sequence. Products and cycling conditions were the same as *RT* products.

Amplification of HIV transcripts from splenocyte RNA: Spleens were thawed and homogenized in 1mL Trizol LS Reagent (Thermo Fisher Scientific). RNA was precipitated and extracted from aqueous phases per manufacturer's protocol. RNA was treated with rDNase I to remove residual DNA (Ambion). Organic phases were saved for DNA extraction. cDNAs were synthesized using the second generation *RT* cDNA primer (4633). From 5' to 3', the second generation primer included a 16 bp T7 sequence, 8 bp randomized sequence (UMID) and 23 bp HIV-specific sequence (AF324493.2: 3305-3327). Viral cDNA was synthesized and column purified as described above. Viral cDNA was amplified in the first round PCR using primers 1922 and 2401 and PfuUltra II Fusion HS Polymerase (95°C for 2 min, followed by 23 cycles of 93°C for 20 sec, 48°C for 20 sec, 72°C for 15 sec and a final extension of 72°C for 10 min). First round products were purified as described above and amplified in a second round PCR using primer 1690 and one of 80 primers with distinct MiSeq barcode identifiers and complementarity to the T7 sequence (Table S1). Second round PCRs used PfuUltra II Fusion HS Polymerase (95°C for 2 min, followed by 25 cycles of 93°C for 20 sec, 48°C for 20 sec, 72°C for 15 sec and a final extension of 72°C for 10 min). Second round products were purified as described above.

Amplification of HIV proviruses from splenocyte DNA: Genomic DNA from splenocytes was extracted from Trizol LS organic phases by an optimized protocol using back extraction buffer (4M guanidium thiocyanate, 50mM sodium citrate, 1M Tris). For amplification of proviruses, a single molecule extension was performed to create single-stranded proviral sequences with UMIDs. Primer 4633 was used to amplify the proviral sequences using PfuUltra II Fusion HS Polymerase (95°C for 2 min, followed by 1 cycle of 93°C for 20 sec, 48°C for 20 sec, 72°C for 15 sec and a final extension of 72°C for 15 min). Products larger than 300 bp were column purified (Zymo Size-Selection). First and second round PCRs were done using the same primers used for splenocyte HIV transcripts. Proviral sequences were amplified using PfuUltra II Fusion HS Polymerase (first round: 95°C for 2 min, followed by 25 cycles of 93°C for 20 sec, 48°C for 20 sec, 72°C for 15 sec and a final extension of 72°C for 10 min; second round: 95°C for 2 min, followed by 30 cycles of 93°C for 20 sec, 48°C for 20 sec, 72°C for 15 sec and a final extension of 72°C for 10 min). Purifications between rounds were done as described above.

MiSeq Library Preparation and MiSeq Instrumentation: Sequencing libraries were run on the Illumina MiSeq to sequence paired-end reads. To prepare libraries, bead-purified PCR products containing Illumina adapters (376 bp *RT* amplicons and 393 bp *vif* amplicons) were quantitated by Qubit dsDNA HS Assay Kit (Invitrogen, Carlsbad, CA). To sequence 250 bp of the RT region, a 6pM final library was run with a 20% spike-in of PhiX Control V3 (Illumina, San Diego, CA) for

2x150 cycles using MiSeq Reagent Kit v2 (250 cycles) or MiSeq Reagent Kit v3 (600 cycles). Custom sequencing primers were used for the forward (1692), index (plasma, 3890; splenocyte transcript/proviral, 4577), and reverse (plasma, 3889; splenocyte transcript/proviral, 4578) reads. To sequence 269 bp of *vif*, a 6pM final library was run with a 15% spike-in of PhiX Control V3 for 2x150 cycles. Custom sequencing primers were used for the forward (4580), index (4577) and reverse (4578) reads.

For HIV *RT* we sequenced a total of 83 HIV RNA samples (69 vRNA, 14 HIV transcripts) from 14 animals with the first generation RT ID primer 4372 and 61 HIV RNA samples (42 vRNA, 19 HIV transcripts) from 20 animals with the second generation RT ID primer 4633. We obtained 11'855'691 high quality pairwise HIV RT reads, which represented 333'258 UMIDs (#4372: 136,241; #4633: 197,017). We also sequenced a total of 21 HIV DNA samples from 21 different mice with the second generation RT ID primer 4633. We obtained 1'878'208 pairwise HIV RT reads that represented 5'547 UMIDs. For HIV *vif*, we sequenced 27 plasma samples and obtained 579'466 reads and identified 6'983 distinct UMIDs.

Bioinformatic Pipeline for Sequence Analyses: Custom Unix and Perl scripts were written to process FASTQ files. First, paired-end reads were merged using the Paired-End reAd mergeR [304]. Reads were aligned to pNL4-3 *RT* or *vif* (AF324493.2) and filtered. Consensus sequences for sequences defined by a distinct UMID were identified using in-house scripts (Figure S3B). Additional custom scripts were written to quantify 3TC resistance mutations at *RT* codon 184 and to compute mutation rates. DNASP v5 polymorphism software was used to calculate the average number of nucleotide differences per site between two sequences, or nucleotide diversity (π) [305]. To identify hypermutants, the Hypermut 2.0 program (https://www.hiv.lanl.gov/content/sequence/HYPERMUT/hypermut.html) was used. Briefly, Fisher's exact test calculated a probability associated with hypermutation based on frequency of G-to-A changes at APOBEC3 target sites (GRD->ARD, R: A or G, D: A, T or G) and non-target sites. If $p<0.05$, the sequence was defined as a hypermutant.

Statistics: Data were analyzed with nonparametric, two-tailed tests, unless the dataset passed normality testing. Significance testing is reported as exact p-values in the text or using asterisks to represent p-values in the figures (i.e. " * "= $p \leq 0.05$; " ** " = $p \leq 0.01$; " *** " = $p \leq 0.001$; " **** "= $p \leq 0.0001$). Results are reported as mean ± standard deviation (SD) and figures depict group mean and SD, unless stated otherwise. All statistics were performed using the built-in analysis packages from GraphPad Prism v7.0 Suite (GraphPad Software).

3.5. Supplementary Material

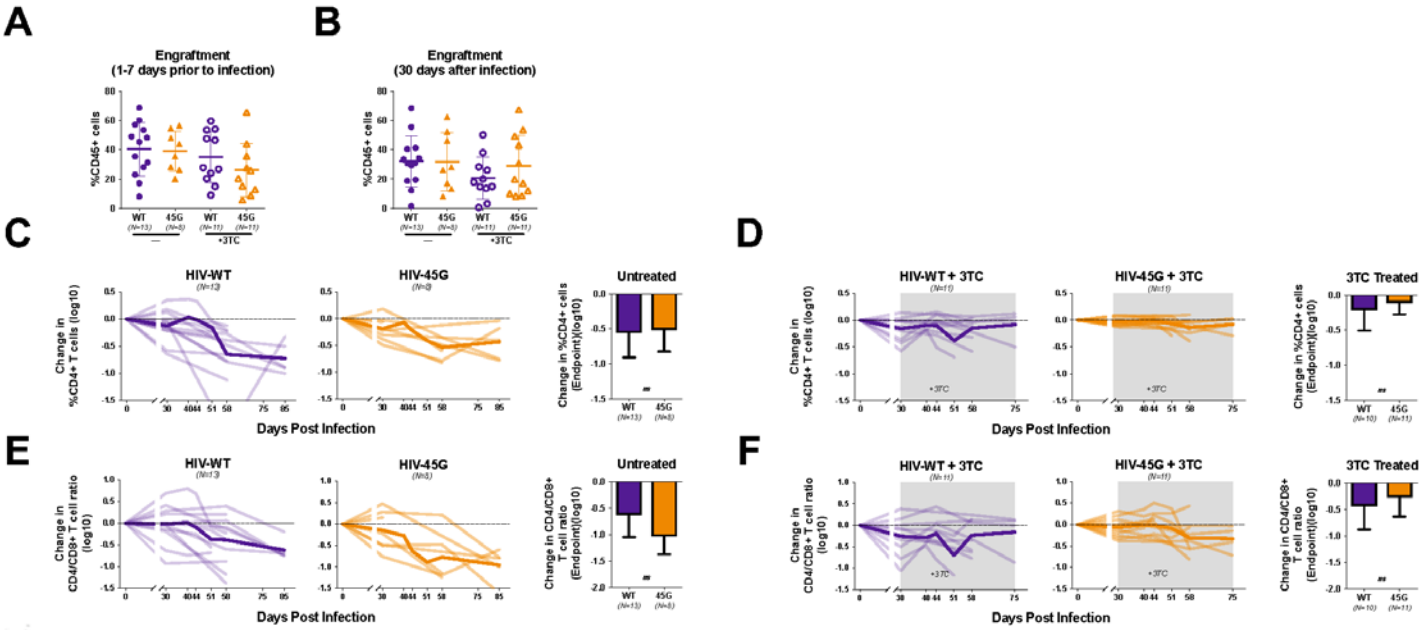


Figure S1. ***Cellular dynamics in hu-mice over time.*** A.-B. *Flow cytometric analyses of human engraftment levels (i.e. %CD45+) at 30 days prior to and after infection in hu-mice intended for treated and untreated groups.* C.-F. *Change in CD4+ T cell frequencies (C, E) and CD4+/CD8+ T cell ratio (D, F) from the day of infection to endpoint in untreated (left-hand side panels) and treated (right-hand side panels) hu-mice. Gray shading represents days on 3TC treatment. Pale lines represent individual mice and bold lines represent group means.*

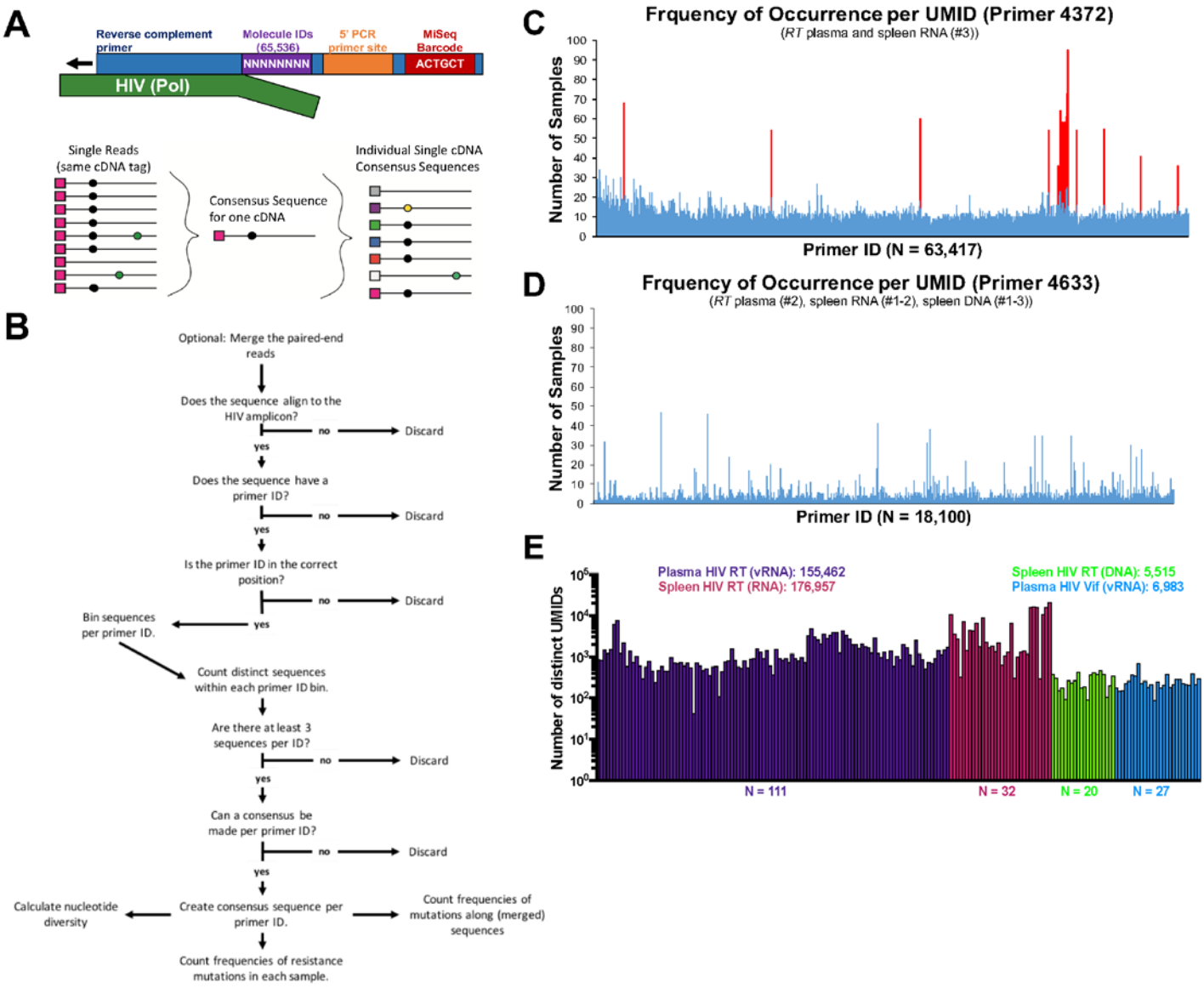


Figure S2. ***Sequencing of HIV in infected hu-mice.*** A. *Schematic depicting the use of a cDNA synthesis primer with a unique molecular tag (UMID), to label and track HIV nucleic acids for sequencing.* B. *Workflow for sequencing HIV nucleic acids.* C.-D. *Number of samples per UMID for the first generation of RT-primer 4372 (C) and second generation of RT-primer 4633 (D). Red bars reflect UMIDs that have only 1 mismatch to the HIV specific sequence and thus have a bias in amplification. These were excluded from downstream analyses.* E. *Number of individual nucleic acid molecules sequenced per sample in plasma and splenocytes for HIV RT and Vif genes.*

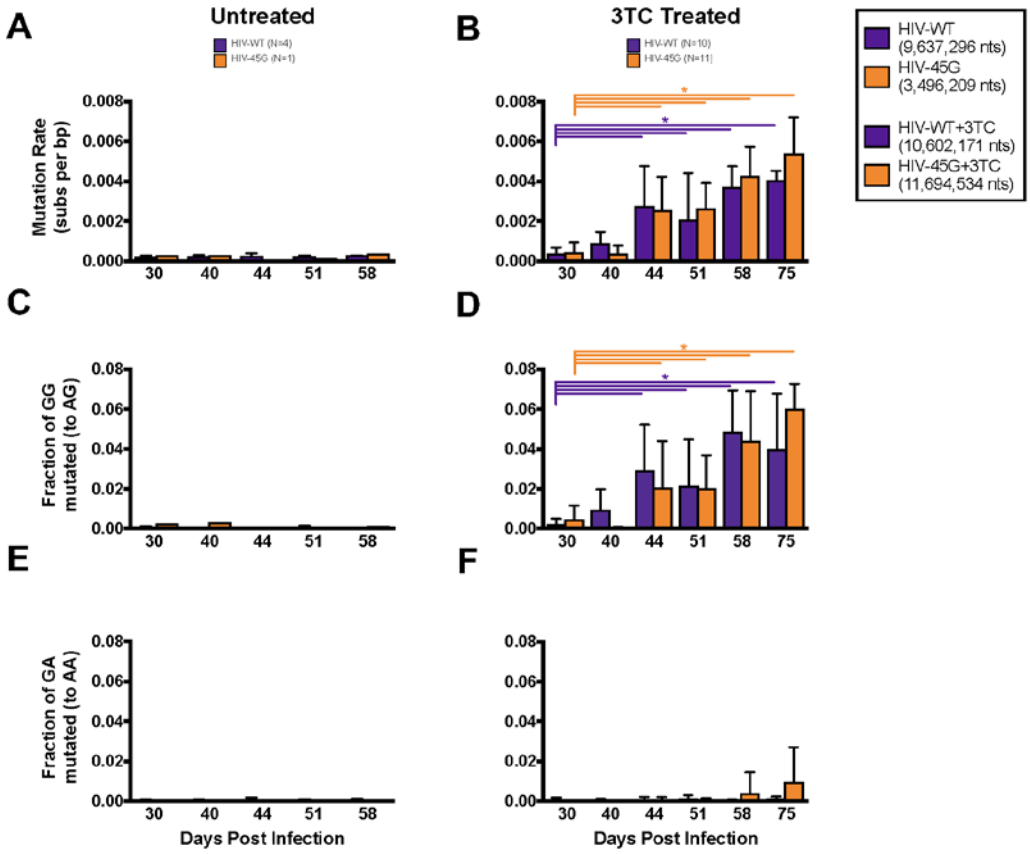


Figure S3. ***Overall and specific G-to-A mutations in plasma quasispecies over time.*** A.-B. *Overall mutation rates in plasma viral RNA of untreated (A) and 3TC treated (B) hu-mice.* C.-D. *Fraction of GG dinucleotides mutated to AG in untreated (C) and treated (D) hu-mice*. E.-F. *Fraction of GA dinucleotides mutated to AA in untreated (E) and treated (F) hu-mice. Mean±SD depicted.*

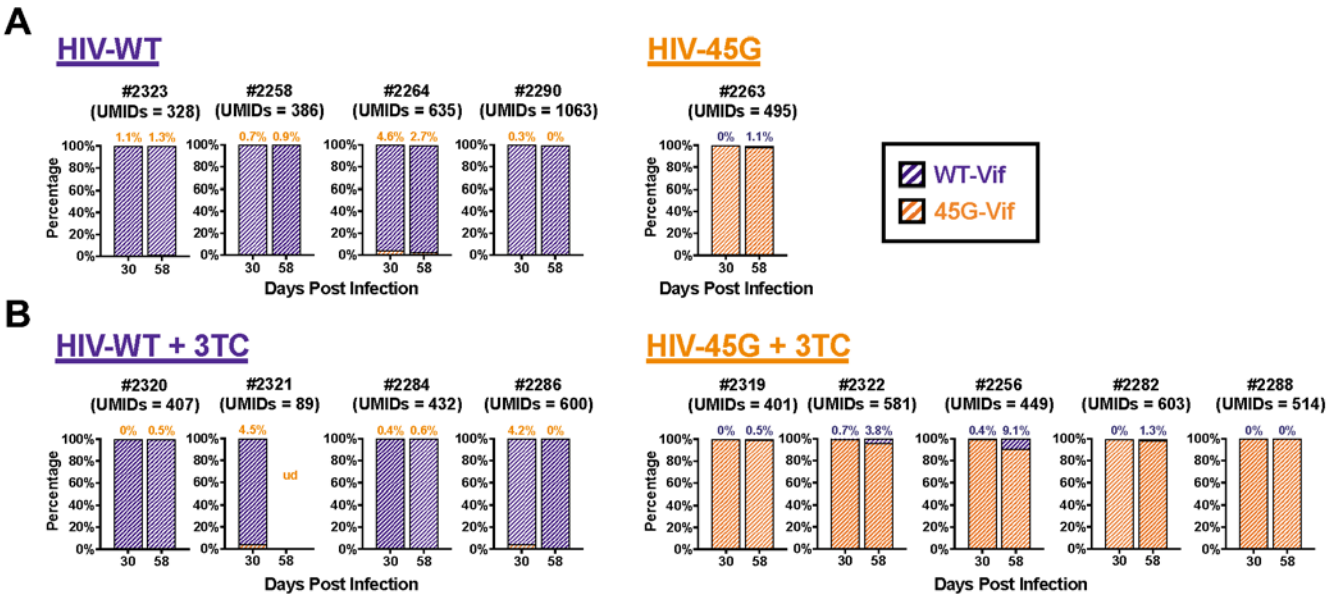


Figure S4. ***Stability of Vif genotypes in plasma over time.*** A.-B. *Frequencies of Vif-E45 (purple) and Vif-45G (orange) genotypes in plasma of untreated (A) and treated (B) hu-mice (from Exp. #3) at 30 and 58 days post infection.*

Reverse Transcription (cDNA synthesis) Primers		
Primer	**HIV Target**	**Sequence (5' to 3')**
4372	*RT (1st gen)*	TTGCCCAATTCAATTTNNNNNNNNATTTCTGTATGTCATTGACAGTC
4373	*Vif*	TAATACGACTCACTATAGGGNNNNNNNNATTTCTTATAGCAGATTCTGA
4633	*RT (2nd gen)*	TTGCCCAATTCAATTTNNNNNNNNATTTCTGTATGTCATTGACAGTC
First Round PCR Primers		
Primer	**HIV Target**	**Sequence (5' to 3')**
1922	*RT (For)*	ATAATACTRCATTTACCATAC
1923	*RT (Rev)*	CTGACTTGCCCAATTCAATT
2402	*Vif (For)*	GTGACATAAAAGTAGTGCCA
2401	*RT/Vif (T7) (Rev)*	TAATACGACTCACTATAGGG
Second Round PCR Primers (Forward Only)		
Primer	**HIV Target**	**Sequence (5' to 3')**
1690	*RT*	AATGATACGGCGACCACCGAGATCTACACTCTTTCGGCCTTTTAGAAAACAAAATC
2403	*Vif*	AATGATACGGCGACCACCGAGATCTACACTCTTTCGACACCATATGTATATTTCAA
Second Round PCR Primers (Reverse with MiSeq Barcodes) (N=74)		
Primer	**HIV Target**	**Sequence (5' to 3')**
2128	*RT*	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTCTGACTTGCCCAATTCAAT
2129	*RT*	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTCTGACTTGCCCAATTCAAT
2130	*RT*	CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTCTGACTTGCCCAATTCAAT
2131	*RT*	CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTCTGACTTGCCCAATTCAAT
2132	*RT*	CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTCTGACTTGCCCAATTCAAT
2133	*RT*	CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTCTGACTTGCCCAATTCAAT
2134	*RT*	CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTCTGACTTGCCCAATTCAAT
2135	*RT*	CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTCTGACTTGCCCAATTCAAT
2136	*RT*	CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTCTGACTTGCCCAATTCAAT
2137	*RT*	CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTCTGACTTGCCCAATTCAAT
2138	*RT*	CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGTCTGACTTGCCCAATTCAAT
2139	*RT*	CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTCTGACTTGCCCAATTCAAT
2140	*RT*	CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTCTGACTTGCCCAATTCAAT
2141	*RT*	CAAGCAGAAGACGGCATACGAGATTTGACTGTGACTGGAGTCTGACTTGCCCAATTCAAT
2142	*RT*	CAAGCAGAAGACGGCATACGAGATGGAACTGTGACTGGAGTCTGACTTGCCCAATTCAAT
2143	*RT*	CAAGCAGAAGACGGCATACGAGATTGACATGTGACTGGAGTCTGACTTGCCCAATTCAAT
2144	*RT*	CAAGCAGAAGACGGCATACGAGATGGACGGGTGACTGGAGTCTGACTTGCCCAATTCAAT
2145	*RT*	CAAGCAGAAGACGGCATACGAGATCTCTACGTGACTGGAGTCTGACTTGCCCAATTCAAT
2146	*RT*	CAAGCAGAAGACGGCATACGAGATGCGGACGTGACTGGAGTCTGACTTGCCCAATTCAAT
2147	*RT*	CAAGCAGAAGACGGCATACGAGATTTTCACGTGACTGGAGTCTGACTTGCCCAATTCAAT
2148	*RT*	CAAGCAGAAGACGGCATACGAGATGGCCACGTGACTGGAGTCTGACTTGCCCAATTCAAT
2149	*RT*	CAAGCAGAAGACGGCATACGAGATCGAAACGTGACTGGAGTCTGACTTGCCCAATTCAAT
2150	*RT*	CAAGCAGAAGACGGCATACGAGATCGTACGGTGACTGGAGTCTGACTTGCCCAATTCAAT
2151	*RT*	CAAGCAGAAGACGGCATACGAGATCCACTCGTGACTGGAGTCTGACTTGCCCAATTCAAT
2152	*RT*	CAAGCAGAAGACGGCATACGAGATGCTACCGTGACTGGAGTCTGACTTGCCCAATTCAAT
2309	*RT*	CAAGCAGAAGACGGCATACGAGATATCAGTGTGACTGGAGTCTGACTTGCCCAATTCAAT
2310	*RT*	CAAGCAGAAGACGGCATACGAGATGCTCATGTGACTGGAGTCTGACTTGCCCAATTCAAT
2311	*RT*	CAAGCAGAAGACGGCATACGAGATAGGAATGTGACTGGAGTCTGACTTGCCCAATTCAAT
2312	*RT*	CAAGCAGAAGACGGCATACGAGATCTTTTGGTGACTGGAGTCTGACTTGCCCAATTCAAT
2313	*RT*	CAAGCAGAAGACGGCATACGAGATTAGTTGGTGACTGGAGTCTGACTTGCCCAATTCAAT
2314	*RT*	CAAGCAGAAGACGGCATACGAGATCCGGTGGTGACTGGAGTCTGACTTGCCCAATTCAAT
2315	*RT*	CAAGCAGAAGACGGCATACGAGATATCGTGGTGACTGGAGTCTGACTTGCCCAATTCAAT
2316	*RT*	CAAGCAGAAGACGGCATACGAGATAAAATGGTGACTGGAGTCTGACTTGCCCAATTCAAT
2317	*RT*	CAAGCAGAAGACGGCATACGAGATATTCCGGTGACTGGAGTCTGACTTGCCCAATTCAAT
2318	*RT*	CAAGCAGAAGACGGCATACGAGATGCTGTAGTGACTGGAGTCTGACTTGCCCAATTCAAT
2319	*RT*	CAAGCAGAAGACGGCATACGAGATGAATGAGTGACTGGAGTCTGACTTGCCCAATTCAAT
2320	*RT*	CAAGCAGAAGACGGCATACGAGATTCGGGAGTGACTGGAGTCTGACTTGCCCAATTCAAT
2321	*RT*	CAAGCAGAAGACGGCATACGAGATCTTCGAGTGACTGGAGTCTGACTTGCCCAATTCAAT
2322	*RT*	CAAGCAGAAGACGGCATACGAGATTGCCGAGTGACTGGAGTCTGACTTGCCCAATTCAAT
2562	*RT*	CAAGCAGAAGACGGCATACGAGATTGCTTCGTGACTGGAGTCTGACTTGCCCAATTCAAT
2563	*RT*	CAAGCAGAAGACGGCATACGAGATACGCGTGTGACTGGAGTCTGACTTGCCCAATTCAAT
2564	*RT*	CAAGCAGAAGACGGCATACGAGATGGGAGCGTGACTGGAGTCTGACTTGCCCAATTCAAT
2565	*RT*	CAAGCAGAAGACGGCATACGAGATCCGACAGTGACTGGAGTCTGACTTGCCCAATTCAAT
2566	*RT*	CAAGCAGAAGACGGCATACGAGATAGTGCAGTGACTGGAGTCTGACTTGCCCAATTCAAT
2567	*RT*	CAAGCAGAAGACGGCATACGAGATACCGCTGTGACTGGAGTCTGACTTGCCCAATTCAAT
2568	*RT*	CAAGCAGAAGACGGCATACGAGATCAAGCAGTGACTGGAGTCTGACTTGCCCAATTCAAT
2569	*RT*	CAAGCAGAAGACGGCATACGAGATTAGCGCGTGACTGGAGTCTGACTTGCCCAATTCAAT
2570	*RT*	CAAGCAGAAGACGGCATACGAGATTACCCTGTGACTGGAGTCTGACTTGCCCAATTCAAT
2571	*RT*	CAAGCAGAAGACGGCATACGAGATAAATACGTGACTGGAGTCTGACTTGCCCAATTCAAT
2572	*RT*	CAAGCAGAAGACGGCATACGAGATCTATCTGTGACTGGAGTCTGACTTGCCCAATTCAAT
2573	*RT*	CAAGCAGAAGACGGCATACGAGATTTATGCGTGACTGGAGTCTGACTTGCCCAATTCAAT
4444	*RT*	CAAGCAGAAGACGGCATACGAGATACATATGTGACTGGAGTCTGACTTGCCCAATTCAAT
4445	*RT*	CAAGCAGAAGACGGCATACGAGATACGTCTGTGACTGGAGTCTGACTTGCCCAATTCAAT
4446	*RT*	CAAGCAGAAGACGGCATACGAGATACTAGCGTGACTGGAGTCTGACTTGCCCAATTCAAT
4447	*RT*	CAAGCAGAAGACGGCATACGAGATACTATAGTGACTGGAGTCTGACTTGCCCAATTCAAT
4448	*RT*	CAAGCAGAAGACGGCATACGAGATAGTCTAGTGACTGGAGTCTGACTTGCCCAATTCAAT
4449	*RT*	CAAGCAGAAGACGGCATACGAGATATTGAGGTGACTGGAGTCTGACTTGCCCAATTCAAT
4450	*RT*	CAAGCAGAAGACGGCATACGAGATATATCAGTGACTGGAGTCTGACTTGCCCAATTCAAT
4451	*RT*	CAAGCAGAAGACGGCATACGAGATATGATTGTGACTGGAGTCTGACTTGCCCAATTCAAT
4452	*RT*	CAAGCAGAAGACGGCATACGAGATATGCTGGTGACTGGAGTCTGACTTGCCCAATTCAAT
4453	*RT*	CAAGCAGAAGACGGCATACGAGATCACGCAGTGACTGGAGTCTGACTTGCCCAATTCAAT
4454	*RT*	CAAGCAGAAGACGGCATACGAGATCATCAGGTGACTGGAGTCTGACTTGCCCAATTCAAT
4455	*RT*	CAAGCAGAAGACGGCATACGAGATCGCAGTGTGACTGGAGTCTGACTTGCCCAATTCAAT
4456	*RT*	CAAGCAGAAGACGGCATACGAGATCATAGCGTGACTGGAGTCTGACTTGCCCAATTCAAT
4457	*RT*	CAAGCAGAAGACGGCATACGAGATCTAGTAGTGACTGGAGTCTGACTTGCCCAATTCAAT
4458	*RT*	CAAGCAGAAGACGGCATACGAGATGACACAGTGACTGGAGTCTGACTTGCCCAATTCAAT
4459	*RT*	CAAGCAGAAGACGGCATACGAGATGTCTACGTGACTGGAGTCTGACTTGCCCAATTCAAT
4460	*RT*	CAAGCAGAAGACGGCATACGAGATGAGATCGTGACTGGAGTCTGACTTGCCCAATTCAAT
4461	*RT*	CAAGCAGAAGACGGCATACGAGATGCACGTGTGACTGGAGTCTGACTTGCCCAATTCAAT
4462	*RT*	CAAGCAGAAGACGGCATACGAGATTGATCAGTGACTGGAGTCTGACTTGCCCAATTCAAT
4463	*RT*	CAAGCAGAAGACGGCATACGAGATTATCTCGTGACTGGAGTCTGACTTGCCCAATTCAAT
4464	*RT*	CAAGCAGAAGACGGCATACGAGATTCGCTCGTGACTGGAGTCTGACTTGCCCAATTCAAT
4465	*RT*	CAAGCAGAAGACGGCATACGAGATTCTCGTGTGACTGGAGTCTGACTTGCCCAATTCAAT
4466	*RT*	CAAGCAGAAGACGGCATACGAGATTGCGATGTGACTGGAGTCTGACTTGCCCAATTCAAT

Second Round PCR Primers (Reverse with MiSeq Barcodes) (N=80)		
Primer	**HIV Target**	**Sequence (5' to 3')**
4634	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTAATACGACTCACTATAGGG
4635	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTAATACGACTCACTATAGGG
4636	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTAATACGACTCACTATAGGG
4637	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTAATACGACTCACTATAGGG
4638	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTAATACGACTCACTATAGGG
4639	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTAATACGACTCACTATAGGG
4640	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTAATACGACTCACTATAGGG
4641	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTAATACGACTCACTATAGGG
4642	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTAATACGACTCACTATAGGG
4643	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGTAATACGACTCACTATAGGG
4644	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTAATACGACTCACTATAGGG
4645	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTAATACGACTCACTATAGGG
4646	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATTTGACTGTGACTGGAGTAATACGACTCACTATAGGG
4647	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATGGAACTGTGACTGGAGTAATACGACTCACTATAGGG
4648	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATTGACATGTGACTGGAGTAATACGACTCACTATAGGG
4649	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATGGACGGGTGACTGGAGTAATACGACTCACTATAGGG
4650	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATCTCTACGTGACTGGAGTAATACGACTCACTATAGGG
4651	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATGCGGACGTGACTGGAGTAATACGACTCACTATAGGG
4652	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATTTTCACGTGACTGGAGTAATACGACTCACTATAGGG
4653	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATGGCCACGTGACTGGAGTAATACGACTCACTATAGGG
4654	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATCGAAACGTGACTGGAGTAATACGACTCACTATAGGG
4655	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATCGTACGGTGACTGGAGTAATACGACTCACTATAGGG
4656	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATCCACTCGTGACTGGAGTAATACGACTCACTATAGGG
4657	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATGCTACCGTGACTGGAGTAATACGACTCACTATAGGG
4658	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATATCAGTGTGACTGGAGTAATACGACTCACTATAGGG
4659	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATGCTCATGTGACTGGAGTAATACGACTCACTATAGGG
4660	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATAGGAATGTGACTGGAGTAATACGACTCACTATAGGG
4661	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATCTTTTGGTGACTGGAGTAATACGACTCACTATAGGG
4662	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATTAGTTGGTGACTGGAGTAATACGACTCACTATAGGG
4663	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATCCGGTGGTGACTGGAGTAATACGACTCACTATAGGG
4664	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATATCGTGGTGACTGGAGTAATACGACTCACTATAGGG
4665	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATAAAATGGTGACTGGAGTAATACGACTCACTATAGGG
4666	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATATTCCGGTGACTGGAGTAATACGACTCACTATAGGG
4667	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATGCTGTAGTGACTGGAGTAATACGACTCACTATAGGG
4668	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATGAATGAGTGACTGGAGTAATACGACTCACTATAGGG
4669	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATTCGGGAGTGACTGGAGTAATACGACTCACTATAGGG
4670	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATCTTCGAGTGACTGGAGTAATACGACTCACTATAGGG
4671	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATTGCTTCGTGACTGGAGTAATACGACTCACTATAGGG
4672	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATACGCGTGTGACTGGAGTAATACGACTCACTATAGGG
4673	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATGGGAGCGTGACTGGAGTAATACGACTCACTATAGGG
4674	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATCCGACAGTGACTGGAGTAATACGACTCACTATAGGG
4675	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATAGTGCAGTGACTGGAGTAATACGACTCACTATAGGG
4676	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATACCGCTGTGACTGGAGTAATACGACTCACTATAGGG
4677	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATTAGCGCGTGACTGGAGTAATACGACTCACTATAGGG
4678	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATAAATACGTGACTGGAGTAATACGACTCACTATAGGG
4679	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATCTATCTGTGACTGGAGTAATACGACTCACTATAGGG
4680	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATTTATGCGTGACTGGAGTAATACGACTCACTATAGGG
4681	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATACATATGTGACTGGAGTAATACGACTCACTATAGGG
4682	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATACGTCTGTGACTGGAGTAATACGACTCACTATAGGG
4683	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATACTAGCGTGACTGGAGTAATACGACTCACTATAGGG
4684	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATACTATAGTGACTGGAGTAATACGACTCACTATAGGG
4685	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATAGTCTAGTGACTGGAGTAATACGACTCACTATAGGG
4686	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATATTGAGGTGACTGGAGTAATACGACTCACTATAGGG
4687	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATATATCAGTGACTGGAGTAATACGACTCACTATAGGG
4688	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATATGATTGTGACTGGAGTAATACGACTCACTATAGGG
4689	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATATGCTGGTGACTGGAGTAATACGACTCACTATAGGG
4690	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATCACGCAGTGACTGGAGTAATACGACTCACTATAGGG
4691	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATCATCAGGTGACTGGAGTAATACGACTCACTATAGGG
4692	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATCGCAGTGTGACTGGAGTAATACGACTCACTATAGGG
4693	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATCATAGCGTGACTGGAGTAATACGACTCACTATAGGG
4694	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATCTAGTAGTGACTGGAGTAATACGACTCACTATAGGG
4695	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATGACACAGTGACTGGAGTAATACGACTCACTATAGGG
4696	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATGTCTACGTGACTGGAGTAATACGACTCACTATAGGG
4697	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATGAGATCGTGACTGGAGTAATACGACTCACTATAGGG
4698	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATGCACGTGTGACTGGAGTAATACGACTCACTATAGGG
4699	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATTGATCAGTGACTGGAGTAATACGACTCACTATAGGG
4700	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATTATCTCGTGACTGGAGTAATACGACTCACTATAGGG
4701	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATTCGCTCGTGACTGGAGTAATACGACTCACTATAGGG
4702	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATTCTCGTGTGACTGGAGTAATACGACTCACTATAGGG
4703	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATTGCGATGTGACTGGAGTAATACGACTCACTATAGGG
4704	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATTGCCGAGTGACTGGAGTAATACGACTCACTATAGGG
4705	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATTACCCTGTGACTGGAGTAATACGACTCACTATAGGG
4722	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATGTTACAGTGACTGGAGTAATACGACTCACTATAGGG
4723	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATGGATTCGTGACTGGAGTAATACGACTCACTATAGGG
4724	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATACAGATGTGACTGGAGTAATACGACTCACTATAGGG
4725	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATTATTGCGTGACTGGAGTAATACGACTCACTATAGGG
4726	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATGTCTATGTGACTGGAGTAATACGACTCACTATAGGG
4727	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATTCTCAGGTGACTGGAGTAATACGACTCACTATAGGG
4728	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATATAGAGGTGACTGGAGTAATACGACTCACTATAGGG
4729	*RT/Vif (T7)*	CAAGCAGAAGACGGCATACGAGATCTACAAGTGACTGGAGTAATACGACTCACTATAGGG
MiSeq Sequencing Primers		
Primer	**Target**	**Sequence (5' to 3')**
1692	*RT For*	ACACTCTTTCCGGCCTTTTAGAAAACAAAATC
3890	*RT* Index	ATTGAATTGGGCAAGTCAGACTCCAGTCAC
3889	*RT* Rev	GTGACTGGAGTCTGACTTGCCCAATTCAAT
4577	*RT/Vif (T7)* Index	CCCTATAGTGAGTCGTATTACTCCAGTCAC
4578	*RT/Vif (T7)* Rev	GTGACTGGAGTAATACGACTCACTATAGGG
4580	*Vif* For	CGAGATCTACACTCTTTCCGACACCATATGTATATTTC

Table S1. Primers for reverse transcription, PCR amplification and Illumina MiSeq Sequencing.

4. VARIA

Audigé et al. BMC Immunology (2017) 18:28
DOI 10.1186/s12865-017-0209-9

BMC Immunology

RESEARCH ARTICLE Open Access

Long-term leukocyte reconstitution in NSG mice transplanted with human cord blood hematopoietic stem and progenitor cells

Annette Audigé[1], Mary-Aude Rochat[1], Duo Li[1], Sandra Ivic[1], Audrey Fahmy[1], Christina K. S. Muller[1], Gustavo Gers-Huber[1], Renier Myburgh[1], Simon Bredl[1], Erika Schlaepfer[1], Alexandra U. Scherrer[1,2], Stefan P. Kuster[1] and Roberto F. Speck[1*]

Abstract

Background: Humanized mice (hu mice) are based on the transplantation of hematopoietic stem and progenitor cells into immunodeficient mice and have become important pre-clinical models for biomedical research. However, data about their hematopoiesis over time are scarce. We therefore characterized leukocyte reconstitution in NSG mice, which were sublethally irradiated and transplanted with human cord blood-derived CD34+ cells at newborn age, longitudinally in peripheral blood and, for more detailed analyses, cross-sectionally in peripheral blood, spleen and bone marrow at different time points.

Results: Human cell chimerism and absolute human cell count decreased between week 16 and 24 in the peripheral blood of hu mice, but were stable thereafter as assessed up to 32 weeks. Human cell chimerism in spleen and bone marrow was maintained over time. Notably, human cell chimerism in peripheral blood and spleen as well as bone marrow positively correlated with each other. Percentage of B cells decreased between week 16 and 24, whereas percentage of T cells increased; subsequently, they levelled off with T cells clearly predominating at week 32. Natural killer cells, monocytes and plasmacytoid dendritic cells (DCs) as well as CD1c + and CD141+ myeloid DCs were all present in hu mice. Proliferative responses of splenic T cells to stimulation were preserved over time. Importantly, the percentage of more primitive hematopoietic stem cells (HSCs) in bone marrow was maintained over time.

Conclusions: Overall, leukocyte reconstitution was maintained up to 32 weeks post-transplantation in our hu NSG model, possibly explained by the maintenance of HSCs in the bone marrow. Notably, we observed great variation in multi-lineage hematopoietic reconstitution in hu mice that needs to be taken into account for the experimental design with hu mice.

Keywords: Hematopoiesis, Hematopoietic stem cells, Cord blood stem cell transplantation, Humanized mice, NSG mice

This study assesses the leukocyte reconstitution in NSG hu-mice, transplanted with cord-blood derived CD34+ cells over time. The main finding are that human engraftment levels were stable for up to 32 weeks in the periphery, spleen and bone marrow, and that this long-term engraftment was supported by the maintenance of primitive hematopoietic stem cells in the bone marrow over time.

Personal contributions: I was involved in the isolation of the cord-blood derived CD34+ cells, generation of the humanized mice and the weekly animal health checks.

Quantitative rather than qualitative differences distinguish the anti-HIV effects of IFN-alpha family members

Erika Schlaepfer[1], Audrey Fahrny[1], Maarja Gruenbach[1], Stefan Kuster[1], Viviana Simon[3], Gideon Schreiber[4], Roberto F. Speck[1]

[1]Division of Infectious Diseases and Hospital Epidemiology, University Hospital of Zurich, 8091 Zurich, Switzerland, [2]Functional Genomics Center, 8057 Zurich, Switzerland, [3]Department of Microbiology, The Global Health and Emerging Pathogens Institute, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA, [4]Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot 76100, Israel

Abstract

The human interferon-alpha (IFN-α) family comprises 13 subtypes. They differ from each other by their binding affinity to IFNARs 1 and 2, which might lead to differential biologic responses. Thus, we interrogated whether IFN-α subtypes also exert distinct anti-human immunodeficiency virus (HIV) activities. We found a dose-dependent response with > 90% inhibition of HIV replication at higher dosages except for the weak IFNAR binder, IFN-α1. Additionally, IFN-derived mutants with substantially reduced IFNAR2 but preserved IFNAR1 binding still inhibited HIV replication efficiently. Notably, the potential to inhibit HIV as well as interferon induced gene induction were more similar between macrophages from the same donor treated with distinct IFN subtypes than that observed between macrophages from different donors treated with the same IFN subtype. These comprehensive studies of dose-relationships demonstrates that IFN subtypes differ in quantitative but not qualitative anti-viral activities. Nonetheless, the question remains whether distinct IFN subtypes such as IFN-α6, -α14 and the IFN-YNS mutant result in differential responses in the complex microenvironment of lymphoid tissues in vivo as compared to the prototypic IFN-α2. The

This study is submitted and assesses the anti-HIV activity of different interferon alpha (IFNa) subtypes, which differ in their binding affinities to IFNARs 1 and 2, in primary cells. The main findings are that the different IFNa subtypes do not differ qualitatively, but rather quantitatively, in their ability to restrict HIV replication. Namely, when high enough doses of each IFNa were used, their anti-HIV effects were comparable. Also, the interferon-stimulated genes (ISGs) response in primary CD4+ T cells after treatment with different IFNa subtypes revealed a similar regulation of ISGs across all conditions, with gene expression patterns clustering with donors rather than the IFNa subtype treatment.

Personal contributions: I was involved in the analysis of interferon-stimulated gene expression data and generation of Figure 4A.

5. CONCLUSIONS AND FUTURE DIRECTIONS

As outlined in the introduction, a major barrier to HIV cure is HIV's ability to establish latent infections, whereby the virus can persist in a silent state for years, despite effective suppression of viral replication with cART. This latent reservoir appears to be seeded throughout the host, in different anatomical compartments and is unaffected by host-immune responses and cART. Yet, upon interruption of cART, viral rebound and progression of HIV infection rapidly occur in almost all HIV-infected patients (even after decades of effective viral suppression by cART), due to the ability of latently HIV-infected cells to revert to a productive state and re-establish a systemic infection [9-11, 13, 16, 67, 306]. Thus, to stop the HIV pandemic, a cure is needed and is likely to require the complete clearing of the latent HIV reservoir. In order to achieve this, we need more insights into HIV latency.

As we lack a ubiquitous and specific means to identify latently infected cells in aviremic HIV-infected patients and access to patient samples is very limited, characterization of the cell types and the mechanisms contributing to HIV latency are largely restricted to *ex vivo* viral outgrowth assays (VOAs) on patient PBMCs and *in vitro* models of latently infected cells. However, HIV infection is a systemic disease, with HIV replication mainly occurring in deep lymphoid tissues where the lymphocyte subsets and immune landscape differ substantially from the peripheral blood [51, 60, 90, 91, 93, 94, 217, 307]. Furthermore, HIV compartmentalization has been observed in patients [52, 53, 82, 197], showing that HIV diversity, replication and reservoirs in given tissues differ according to the anatomical compartment. Together this suggests that studying the peripheral blood HIV reservoir in patients is not very reflective of latently infected cells in other anatomical compartments. Additionally, *in vitro* models of latently infected cells do not fully recapitulate the *ex vivo* response characteristics of reservoir cells in patients [145]. It is noteworthy that the *ex vivo* response of PBMCs from aviremic patients to different LRAs also greatly varies between donors [242]. The lack of a representative model of systemic HIV latency greatly impedes efforts in HIV cure, with the implication that so far, no pharmacological intervention has been able to block, or even significantly delay, viral rebound in patients after cART interruption [142]. Strikingly, even though IL-2 alone or in combination with IFN-gamma, induced a reduction in the frequency of latently infected PBMCs in patients, viral rebound still rapidly occurred after cART interruption [134]. Taken together, these findings highlight how crucial it is to gain more insight into genuine HIV reservoirs seeded beyond the blood compartment. Hereby I do not seek to devalue *in vitro* models of HIV latency and the key insights these have provided, but rather my intention is to highlight that as HIV seeds a systemic and heterogeneous reservoir, systemic investigations and combinatorial latency reversing approaches will be required if we hope to cure HIV-infected patients.

In the first part of this thesis, we endeavored to develop an *in vivo* model to study HIV latency systemically. We used the NSG hu-mouse model as a basis for our model, as hu-mice are well-established and accepted models for studying HIV pathogenesis and persistence [160, 161]. In addition, as the HSPCs used to reconstitute the mice can be genetically modified prior to transplantation [205, 208], hu-mice provide a uniquely mutable system to work with. To be able to specifically identify latently infected cells in our HIV reporter hu-mouse model, we incorporated an HIV-inducible RFP reporter transgene (i.e. HR4lox transgene) into the NSG hu-mouse model. Specifically, the transgene has two parts: a cassette containing a CMV promoter and floxed (inverted) RFP, followed by a PGK promoter and GFP cassette, which means transgenic cells are constitutively GFP+ and only constitutively express RFP after two Cre-mediated recombination events take place at the lox sites flanking the RFP gene. By using a HIV variant that encodes a CMV promoter-Cre cassette (i.e. HOCC), Cre-lox recombination becomes dependent on HIV infection. Thus in HOCC-infected HR4lox hu-mice, GFP+RFP+ cells comprise a mixture of productively and latently infected cells, whereas in HOCC-infected but cART-suppressed HR4lox hu-mice (i.e. latency scenario), GFP+RFP+ cells identify latently infected cells. Although our HIV reporter hu-mouse model allowed us to isolate infected cells in the productive infection scenario, in the context of cART the resolution of the system was too low to unambiguously identify and sort reservoir cells. Nevertheless, we are convinced the *in vivo* model developed herein is a promising approach for obtaining genuine latently HIV-infected cell from diverse anatomical compartments and is therefore a good basis for further optimization. The two main issues with our model appear to be: i) the use of the CMV promoter for gene expression, which may be silenced by methylation *in vivo*; and ii) the high barrier to effective Cre-lox recombination imposed by the lox site design in the HR4lox transgene. In the discussion section of chapter 2, I elaborate on modifications that might improve the resolution of this HIV latency model. In addition, the use of an alternative reporter gene to RFP may help reduce attenuation of the reporter signal, which currently can arise due to spectral overlap with the GFP signal in transgenic cells and tissue autofluorescence. Once we have improved our HIV reporter hu-mouse model, our ultimate goal is to characterize the cell types harboring silent HIV in as many tissues as possible and carryout single cell RNA sequencing, which may reveal new ways to specifically target latency maintenance and reversion *in vivo*.

Another critical barrier to HIV eradication is HIV's ability to rapidly diversify after transmission. This key attribute of HIV means that it exists in its host as a pool of genetically diverse and continually evolving quasispecies that undermine the development of an effective anti-HIV host immune response and anti-HIV vaccine, and supports the emergence of drug-resistant strains [173, 261]. Although high *in vivo* rates of replication and RT-mediated mutagenesis, as well as recombination between co-packaged genomes are mechanism explicitly used by HIV for diversification, whether APOBEC3-mediated mutagenesis contributes to HIV-1 diversification *in vivo* remains highly

debated. Thus, in the second part of this thesis, we also explored the contribution of APOBEC3G (A3G)-induced mutagenesis to viral diversification and adaption using the NSG hu-mouse model. A3G is a cytidine deaminase that is part of the human innate immune system and potently restricts HIV replication (in the absence of the viral accessory protein Vif) by catalyzing extensive G-to-A substitutions in the nascent viral cDNA during reverse transcription [184]. As A3G's mutagenic activity on HIV often introduces premature stop codons or missense mutations in the viral genome [47, 114], some studies support the notion that A3G-mediated mutagenesis only results in defective (hypermutated) proviruses [176, 193, 194, 270]. On the other hand, A3G has the potential to contribute to HIV sequence diversification if the level of genome editing is sublethal or if hypermutated viral genomes are rescued by recombination with intact viral sequences [179, 192, 267]. Elucidating whether A3G plays a role in viral evolution is relevant to understanding the emergence of drug-resistances, especially as patients carrying HIV mutants with partially defective vif alleles are more likely to fail ART [189, 190, 265, 266]). In addition, it is also important for the design and development of novel antiviral treatments aimed at blocking Vif-APOBEC3 interactions (reviewed in [308]), as partial inhibition of Vif may turn out to be counterproductive and favor viral diversification and escape.

In this study, we show that in hu-mice infected with a HIV clone encoding a Vif mutant with suboptimal A3G-neutralizing capabilities, the splenic proviral quasispecies were more diversified than in hu-mice infected with WT HIV. This increased proviral diversity was characterized by footprints of A3G-mutagenesis and a portion of mutated proviral variants were also represented as cell-associated viral transcripts in the spleen and cell-free virions in the plasma. In addition, the Vif mutant virus's adaptation to monotherapy with the reverse transcriptase inhibitor lamivudine (3TC) was superior compared to WT HIV. This enhanced fitness phenotype was uniquely associated with a preferential emergence of the 3TC-drug resistance mutation M184I, which results from a G-to-A substitution in an A3G-favored dinucleotide context, over other 3TC-resistance mutations. Taken together, our data support that sublethal A3G-mutagenesis can occur *in vivo* and that A3G can shape the proviral landscape, thereby accelerating the development of drug resistances.

I believe that the use of hu-mice and isogenic HIV clones differing in their ability to counteract A3G are key strengths of this study. As HIV diversification is shaped by the interplay between the virus and the host immune system, it is crucial to study viral diversification in a system that captures the complexity of the human immune system, especially considering that the expression and activity of A3G is not homogenous throughout tissues and cell types [309, 310]. Indeed, we initially found that the Vif mutant HIV replicated as well as the WT virus *in vitro*, indicating that *in vitro* assays do not reflect accurately the APOBEC3-Vif axis. Moreover, as it is difficult to unequivocally assess whether G-to-A mutations arise from RT-error or APOBEC3-activity, the use of isogenic HIV clones is an interesting approach to dissect the contribution of A3G over RT-error to HIV diversification.

Specifically, the partially defective Vif mutant HIV used herein contains a single point mutation in the A3G-binding domain of Vif. This amino acid has not been specifically associated with Vif functions other than A3G-binding and is not part of any of the other HIV gene products [280, 281]. Thus, difference observed in hu-mice infected with either of the two viruses should be the result of partially unchecked A3G-activity. Moreover, the use of UMIDs in our sequencing methods is another key strength of this work, as it allowed us to sequence viral quasispecies in an unbiased manner and with high-resolution.

Overall, in this thesis, we sought to address key characteristics of HIV that preclude its eradication, namely HIV latency and HIV diversification. To this end, we made extensive use of hu-mice, which are powerful tools for study human diseases.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Prof. Roberto Speck, for giving me the opportunity to conduct research in his laboratory and work on many exciting and diverse projects. I greatly appreciate the freedom he gave me over the years to explore my own ideas and that he always pushed me to expand my horizons and attend conferences and workshops, as well as his unwavering enthusiasm.

I would like to thank my thesis committee members Prof. Maries van den Broek, Prof. Michael Hottiger, Prof. Christian Münz and Prof. Monsef Benkirane for their great advice and constructive criticisms, as well as the rich discussions during our meetings. I would like to thank our collaborators Prof. Monsef Benkirane and Prof. Viviana Simon for the opportunity to share expertise and particularly, I would like to extend my gratitude to Dr. Gaël Petitjean for the great teamwork and interesting discussions.

This thesis would not have been possible without the help of my colleagues and therefore I would like to thank all the members of the Speck lab for their continuous support and advice over the years. A special thank you to Erika Schläpfer for always offering to help and providing support and Christina Müller for the good teamwork and fun times at conferences. I would also like to acknowledge the outstanding competency of the flow cytometry core facility of the University of Zurich and friendliness of their staff.

Last, but no least, I would like to thank David for supporting and comforting me throughout this process and patiently listening to my rants about my research, as well as my parents for always encouraging me to pursue my goals.

CURRICULUM VITAE

FAHRNY Audrey

Date of Birth: 27.12.1989

Le Locle, NE

Nationalities: Swiss and French

Education

since Jan. 2014	**Doctoral thesis** Prof. Dr. med. Roberto Speck, *University Hospital Zurich, CH* *Department of Infectious Diseases and Hospital Epidemiology* *"Exploring HIV Latency and Dissecting the Contribution of APOBEC3G to HIV Evolution using Humanized Mice Models"*
2011 - 2012	**Master of Science in Biomedical Engineering (Distinction)** *Imperial College London, UK* Principal subjects: Biomaterials and tissue engineering, radiotherapy physics and radiobiology, computational neuroscience, biomedical imaging, biomechanics, physiological monitoring, health & safety & regulatory requirements
2011 - 2012	**Master Thesis** Prof. R. Ethier *Imperial College London, UK* *Department of Biomedical Engineering* "Strangling neurons: how much can they take?"
2008 - 2011	**Bachelor of Science in Life Sciences and technology** *École Polytechnique Fédérale de Lausanne (EPFL), CH*
2011	**Bachelor Thesis** Prof. O. Blanke *EPFL, CH,* *Brain Mind Institute – Neurosciences, Laboratory of Cognitive Neuroscience* "Investigating the link between illusory ownership with a virtual body and interoceptive awareness."
2006 - 2008	**International Baccalaureate** *Munich International School (MIS), DE*

REFERENCES

1. Haverkos, H.W. and J.W. Curran, *The current outbreak of Kaposi's sarcoma and opportunistic infections.* CA Cancer J Clin, 1982. **32**(6): p. 330-9.
2. Gallo, R.C., et al., *Frequent Detection and Isolation of Cytopathic Retroviruses (Htlv-Iii) from Patients with Aids and at Risk for Aids.* Science, 1984. **224**(4648): p. 500-503.
3. Popovic, M., et al., *Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS.* Science, 1984. **224**(4648): p. 497-500.
4. WHO. *HIV/AIDS Fact sheet*. 2017 November 2017 10.04.2018]; Available from: http://www.who.int/mediacentre/factsheets/fs360/en/.
5. Sharp, P.M. and B.H. Hahn, *Origins of HIV and the AIDS pandemic.* Cold Spring Harb Perspect Med, 2011. **1**(1): p. a006841.
6. Douek, D.C., M. Roederer, and R.A. Koup, *Emerging concepts in the immunopathogenesis of AIDS.* Annu Rev Med, 2009. **60**: p. 471-84.
7. Low, A., et al., *Incidence of Opportunistic Infections and the Impact of Antiretroviral Therapy Among HIV-Infected Adults in Low- and Middle-Income Countries: A Systematic Review and Meta-analysis.* Clin Infect Dis, 2016. **62**(12): p. 1595-1603.
8. Palella, F.J., Jr., et al., *Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. HIV Outpatient Study Investigators.* N Engl J Med, 1998. **338**(13): p. 853-60.
9. Chun, T.W., et al., *In vivo fate of HIV-1-infected T cells: quantitative analysis of the transition to stable latency.* Nat Med, 1995. **1**(12): p. 1284-90.
10. Chun, T.W., et al., *Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy.* Proc Natl Acad Sci U S A, 1997. **94**(24): p. 13193-7.
11. Finzi, D., *Identification of a Reservoir for HIV-1 in Patients on Highly Active Antiretroviral Therapy.* Science, 1997. **278**(5341): p. 1295-1300.
12. Whitney, J.B., et al., *Rapid seeding of the viral reservoir prior to SIV viraemia in rhesus monkeys.* Nature, 2014. **512**(7512): p. 74-7.
13. Chun, T.W., et al., *Early establishment of a pool of latently infected, resting CD4(+) T cells during primary HIV-1 infection.* Proc Natl Acad Sci U S A, 1998. **95**(15): p. 8869-73.
14. Finzi, D., et al., *Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy.* Nat Med, 1999. **5**(5): p. 512-7.
15. Joos, B., et al., *HIV rebounds from latently infected cells, rather than from continuing low-level replication.* Proc Natl Acad Sci U S A, 2008. **105**(43): p. 16725-30.
16. Chun, T.W., et al., *Re-emergence of HIV after stopping therapy.* Nature, 1999. **401**(6756): p. 874-5.
17. Richman, D.D., et al., *The challenge of finding a cure for HIV infection.* Science, 2009. **323**(5919): p. 1304-7.
18. Cihlar, T. and M. Fordyce, *Current status and prospects of HIV treatment.* Curr Opin Virol, 2016. **18**: p. 50-6.
19. WHO. *HIV/AIDS Data and Statistics*. 2016 29 November 2016 10.04.2018]; Available from: http://www.who.int/hiv/data/en/.
20. Mansky, L.M. and H.M. Temin, *Lower in-Vivo Mutation-Rate of Human-Immunodeficiency-Virus Type-1 Than That Predicted from the Fidelity of Purified Reverse-Transcriptase.* Journal of Virology, 1995. **69**(8): p. 5087-5094.
21. Perelson, A.S., et al., *HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time.* Science, 1996. **271**(5255): p. 1582-6.
22. Robertson, D.L., et al., *Recombination in HIV-1.* Nature, 1995. **374**(6518): p. 124-6.
23. Jetzt, A.E., et al., *High rate of recombination throughout the human immunodeficiency virus type 1 genome.* J Virol, 2000. **74**(3): p. 1234-40.
24. Domingo, E., J. Sheldon, and C. Perales, *Viral quasispecies evolution.* Microbiol Mol Biol Rev, 2012. **76**(2): p. 159-216.
25. Coffin, J. and R. Swanstrom, *HIV pathogenesis: dynamics and genetics of viral populations and infected cells.* Cold Spring Harb Perspect Med, 2013. **3**(1): p. a012526.

26. Gao, Y., P.F. McKay, and J.F.S. Mann, *Advances in HIV-1 Vaccine Development.* Viruses, 2018. **10**(4).
27. Spragg, C., H. De Silva Feelixge, and K.R. Jerome, *Cell and gene therapy strategies to eradicate HIV reservoirs.* Curr Opin HIV AIDS, 2016. **11**(4): p. 442-9.
28. Spivak, A.M. and V. Planelles, *HIV-1 Eradication: Early Trials (and Tribulations).* Trends Mol Med, 2016. **22**(1): p. 10-27.
29. Spivak, A.M. and V. Planelles, *Novel Latency Reversal Agents for HIV-1 Cure.* Annu Rev Med, 2018. **69**: p. 421-436.
30. Faust, T.B., et al., *Making Sense of Multifunctional Proteins: Human Immunodeficiency Virus Type 1 Accessory and Regulatory Proteins and Connections to Transcription.* Annu Rev Virol, 2017. **4**(1): p. 241-260.
31. Rosa, A., et al., *HIV-1 Nef promotes infection by excluding SERINC5 from virion incorporation.* Nature, 2015. **526**(7572): p. 212-7.
32. Flint, S.J., et al., *Principles of Virology*. 3 ed. Vol. I. 2008: ASM Press.
33. Roebuck, K.A. and M. Saifuddin, *Regulation of HIV-1 transcription.* Gene Expr, 1999. **8**(2): p. 67-84.
34. Bleul, C.C., et al., *The HIV coreceptors CXCR4 and CCR5 are differentially expressed and regulated on human T lymphocytes.* Proc Natl Acad Sci U S A, 1997. **94**(5): p. 1925-30.
35. Pollakis, G. and W.A. Paxton, *Use of (alternative) coreceptors for HIV entry.* Curr Opin HIV AIDS, 2012. **7**(5): p. 440-9.
36. Schröder, A.R.W., et al., *HIV-1 Integration in the Human Genome Favors Active Genes and Local Hotspots.* Cell, 2002. **110**(4): p. 521-529.
37. Stevenson, M., et al., *Hiv-1 Replication Is Controlled at the Level of T-Cell Activation and Proviral Integration.* Embo Journal, 1990. **9**(5): p. 1551-1560.
38. Vaishnav, Y.N. and F. Wong-Staal, *The biochemistry of AIDS.* Annu Rev Biochem, 1991. **60**: p. 577-630.
39. Doitsh, G. and W.C. Greene, *Dissecting How CD4 T Cells Are Lost During HIV Infection.* Cell Host Microbe, 2016. **19**(3): p. 280-91.
40. Borrow, P., et al., *Antiviral pressure exerted by HIV-l-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus.* Nature Medicine, 1997. **3**(2): p. 205-211.
41. Ho, D.D., et al., *Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection.* Nature, 1995. **373**(6510): p. 123-6.
42. Wei, X., et al., *Viral dynamics in human immunodeficiency virus type 1 infection.* Nature, 1995. **373**(6510): p. 117-22.
43. Palmer, S., *Advances in detection and monitoring of plasma viremia in HIV-infected individuals receiving antiretroviral therapy.* Curr Opin HIV AIDS, 2013. **8**(2): p. 87-92.
44. Schupbach, J., *Measurement of HIV-1 p24 Antigen by Signal-amplification-boosted ELISA of Heat-denatured Plasma is a Simple and Inexpensive Alternative to Tests for Viral RNA.* AIDS Reviews, 2002. **4**(2): p. 83-92.
45. Alvarez, M., et al., *Improving Clinical Laboratory Efficiency: Introduction of Systems for the Diagnosis and Monitoring of HIV Infection.* Open Virol J, 2012. **6**: p. 135-43.
46. Woodham, A.W., et al., *Human Immunodeficiency Virus Immune Cell Receptors, Coreceptors, and Cofactors: Implications for Prevention and Treatment.* AIDS Patient Care STDS, 2016. **30**(7): p. 291-306.
47. Bruner, K.M., et al., *Defective proviruses rapidly accumulate during acute HIV-1 infection.* Nat Med, 2016. **22**(9): p. 1043-9.
48. Chun, T.W., et al., *Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection.* Nature, 1997. **387**(6629): p. 183-8.
49. Baxter, A.E., U. O'Doherty, and D.E. Kaufmann, *Beyond the replication-competent HIV reservoir: transcription and translation-competent reservoirs.* Retrovirology, 2018. **15**(1): p. 18.
50. Sloan, R.D. and M.A. Wainberg, *The role of unintegrated DNA in HIV infection.* Retrovirology, 2011. **8**: p. 52.
51. Pantaleo, G., et al., *Lymphoid Organs Function as Major Reservoirs for Human-Immunodeficiency-Virus.* Proceedings of the National Academy of Sciences of the United States of America, 1991. **88**(21): p. 9838-9842.

52. Blackard, J.T., *HIV compartmentalization: a review on a clinically important phenomenon.* Curr HIV Res, 2012. **10**(2): p. 133-42.
53. Fulcher, J.A., et al., *Compartmentalization of human immunodeficiency virus type 1 between blood monocytes and CD4+ T cells during infection.* J Virol, 2004. **78**(15): p. 7883-93.
54. Farber, D.L., N.A. Yudanin, and N.P. Restifo, *Human memory T cells: generation, compartmentalization and homeostasis.* Nat Rev Immunol, 2014. **14**(1): p. 24-35.
55. Nabel, G. and D. Baltimore, *An inducible transcription factor activates expression of human immunodeficiency virus in T cells.* Nature, 1987. **326**(6114): p. 711-3.
56. Althaus, C.L., et al., *Quantifying the turnover of transcriptional subclasses of HIV-1-infected cells.* PLoS Comput Biol, 2014. **10**(10): p. e1003871.
57. Shan, L., et al., *Transcriptional Reprogramming during Effector-to-Memory Transition Renders CD4(+) T Cells Permissive for Latent HIV-1 Infection.* Immunity, 2017. **47**(4): p. 766-775 e3.
58. Baxter, A.E., et al., *Single-Cell Characterization of Viral Translation-Competent Reservoirs in HIV-Infected Individuals.* Cell Host Microbe, 2016. **20**(3): p. 368-380.
59. Grau-Exposito, J., et al., *A Novel Single-Cell FISH-Flow Assay Identifies Effector Memory CD4(+) T cells as a Major Niche for HIV-1 Transcription in HIV-Infected Patients.* MBio, 2017. **8**(4).
60. Perreau, M., et al., *Follicular helper T cells serve as the major CD4 T cell compartment for HIV-1 infection, replication, and production.* J Exp Med, 2013. **210**(1): p. 143-56.
61. Brenchley, J.M., et al., *Differential Th17 CD4 T-cell depletion in pathogenic and nonpathogenic lentiviral infections.* Blood, 2008. **112**(7): p. 2826-35.
62. Zack, J.A., et al., *Hiv-1 Entry into Quiescent Primary Lymphocytes - Molecular Analysis Reveals a Labile, Latent Viral Structure.* Cell, 1990. **61**(2): p. 213-222.
63. Pan, X., et al., *Restrictions to HIV-1 replication in resting CD4+ T lymphocytes.* Cell Res, 2013. **23**(7): p. 876-85.
64. Clayton, K.L., et al., *HIV Infection of Macrophages: Implications for Pathogenesis and Cure.* Pathog Immun, 2017. **2**(2): p. 179-192.
65. Sattentau, Q.J. and M. Stevenson, *Macrophages and HIV-1: An Unhealthy Constellation.* Cell Host Microbe, 2016. **19**(3): p. 304-10.
66. Perelson, A.S., et al., *Decay characteristics of HIV-1-infected compartments during combination therapy.* Nature, 1997. **387**(6629): p. 188-91.
67. Shen, L. and R.F. Siliciano, *Viral reservoirs, residual viremia, and the potential of highly active antiretroviral therapy to eradicate HIV infection.* J Allergy Clin Immunol, 2008. **122**(1): p. 22-8.
68. Bui, J.K., et al., *Ex vivo activation of CD4+ T-cells from donors on suppressive ART can lead to sustained production of infectious HIV-1 from a subset of infected cells.* PLoS Pathog, 2017. **13**(2): p. e1006230.
69. Deng, K., et al., *Broad CTL response is required to clear latent HIV-1 due to dominance of escape mutations.* Nature, 2015. **517**(7534): p. 381-5.
70. Migueles, S.A. and M. Connors, *Success and failure of the cellular immune response against HIV-1.* Nat Immunol, 2015. **16**(6): p. 563-70.
71. Simon, V., N. Bloch, and N.R. Landau, *Intrinsic host restrictions to HIV-1 and mechanisms of viral escape.* Nat Immunol, 2015. **16**(6): p. 546-553.
72. Iglesias-Ussel, M., et al., *High levels of CD2 expression identify HIV-1 latently infected resting memory CD4+ T cells in virally suppressed subjects.* J Virol, 2013. **87**(16): p. 9148-58.
73. Descours, B., et al., *CD32a is a marker of a CD4 T-cell HIV reservoir harbouring replication-competent proviruses.* Nature, 2017. **543**(7646): p. 564-567.
74. Siliciano, J.D., et al., *Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells.* Nat Med, 2003. **9**(6): p. 727-8.
75. Gutierrez, C., et al., *Bryostatin-1 for latent virus reactivation in HIV-infected patients on antiretroviral therapy.* AIDS, 2016. **30**(9): p. 1385-92.
76. Sogaard, O.S., et al., *The Depsipeptide Romidepsin Reverses HIV-1 Latency In Vivo.* PLoS Pathog, 2015. **11**(9): p. e1005142.

77. Archin, N.M., et al., *Interval dosing with the HDAC inhibitor vorinostat effectively reverses HIV latency.* J Clin Invest, 2017. **127**(8): p. 3126-3135.
78. Ananworanich, J., B. McSteen, and M.L. Robb, *Broadly neutralizing antibody and the HIV reservoir in acute HIV infection: a strategy toward HIV remission?* Curr Opin HIV AIDS, 2015. **10**(3): p. 198-206.
79. Dinoso, J.B., et al., *Treatment intensification does not reduce residual HIV-1 viremia in patients on highly active antiretroviral therapy.* Proc Natl Acad Sci U S A, 2009. **106**(23): p. 9403-8.
80. Arainga, M., et al., *A mature macrophage is a principal HIV-1 cellular reservoir in humanized mice after treatment with long acting antiretroviral therapy.* Retrovirology, 2017. **14**(1): p. 17.
81. Embretson, J., et al., *Massive covert infection of helper T lymphocytes and macrophages by HIV during the incubation period of AIDS.* Nature, 1993. **362**(6418): p. 359-62.
82. Wong, J.K. and S.A. Yukl, *Tissue reservoirs of HIV.* Curr Opin HIV AIDS, 2016. **11**(4): p. 362-70.
83. Sebastian, N.T. and K.L. Collins, *Targeting HIV latency: resting memory T cells, hematopoietic progenitor cells and future directions.* Expert Rev Anti Infect Ther, 2014. **12**(10): p. 1187-201.
84. Coleman, C.M. and L. Wu, *HIV interactions with monocytes and dendritic cells: viral latency and reservoirs.* Retrovirology, 2009. **6**: p. 51.
85. Lee, G.Q., et al., *Clonal expansion of genome-intact HIV-1 in functionally polarized Th1 CD4+ T cells.* J Clin Invest, 2017. **127**(7): p. 2689-2696.
86. Hosmane, N.N., et al., *Proliferation of latently infected CD4(+) T cells carrying replication-competent HIV-1: Potential role in latent reservoir dynamics.* J Exp Med, 2017. **214**(4): p. 959-972.
87. von Stockenstrom, S., et al., *Longitudinal Genetic Characterization Reveals That Cell Proliferation Maintains a Persistent HIV Type 1 DNA Pool During Effective HIV Therapy.* J Infect Dis, 2015. **212**(4): p. 596-607.
88. Bailey, J.R., et al., *Residual human immunodeficiency virus type 1 viremia in some patients on antiretroviral therapy is dominated by a small number of invariant clones rarely found in circulating CD4+ T cells.* J Virol, 2006. **80**(13): p. 6441-57.
89. Gratton, S., et al., *Highly restricted spread of HIV-1 and multiply infected cells within splenic germinal centers.* Proc Natl Acad Sci U S A, 2000. **97**(26): p. 14566-71.
90. Haase, A.T., *Population biology of HIV-1 infection: viral and CD4+ T cell demographics and dynamics in lymphatic tissues.* Annu Rev Immunol, 1999. **17**: p. 625-56.
91. Kinter, A., et al., *Productive HIV infection of resting CD4+ T cells: role of lymphoid tissue microenvironment and effect of immunomodulating agents.* AIDS Res Hum Retroviruses, 2003. **19**(10): p. 847-56.
92. Pantaleo, G., et al., *HIV infection is active and progressive in lymphoid tissue during the clinically latent stage of disease.* Nature, 1993. **362**(6418): p. 355-8.
93. Thome, J.J., et al., *Early-life compartmentalization of human T cell differentiation and regulatory function in mucosal and lymphoid tissues.* Nat Med, 2016. **22**(1): p. 72-7.
94. Thome, J.J., et al., *Spatial map of human T cell compartmentalization and maintenance over decades of life.* Cell, 2014. **159**(4): p. 814-28.
95. Bailey, J., et al., *Mechanisms of HIV-1 escape from immune responses and antiretroviral drugs.* Curr Opin Immunol, 2004. **16**(4): p. 470-6.
96. Razooky, B.S., et al., *A hardwired HIV latency program.* Cell, 2015. **160**(5): p. 990-1001.
97. Rouzine, I.M., A.D. Weinberger, and L.S. Weinberger, *An evolutionary role for HIV latency in enhancing viral transmission.* Cell, 2015. **160**(5): p. 1002-12.
98. Haase, A.T., *Early events in sexual transmission of HIV and SIV and opportunities for interventions.* Annu Rev Med, 2011. **62**: p. 127-39.
99. Gray, R.H., et al., *Probability of HIV-1 transmission per coital act in monogamous, heterosexual, HIV-1-discordant couples in Rakai, Uganda.* The Lancet, 2001. **357**(9263): p. 1149-1153.
100. Siliciano, R.F. and W.C. Greene, *HIV latency.* Cold Spring Harb Perspect Med, 2011. **1**(1): p. a007096.

101. Ostrowski, M.A., et al., *Both memory and CD45RA(+)/CD62L(+) naive CD4(+) T cells are infected in human immunodeficiency virus type 1-infected individuals.* Journal of Virology, 1999. **73**(8): p. 6430-6435.
102. Baldauf, H.M., et al., *SAMHD1 restricts HIV-1 infection in resting CD4(+) T cells.* Nat Med, 2012. **18**(11): p. 1682-7.
103. Swiggard, W.J., et al., *Human immunodeficiency virus type 1 can establish latent infection in resting CD4+ T cells in the absence of activating stimuli.* J Virol, 2005. **79**(22): p. 14179-88.
104. Saleh, S., et al., *CCR7 ligands CCL19 and CCL21 increase permissiveness of resting memory CD4+ T cells to HIV-1 infection: a novel model of HIV-1 latency.* Blood, 2007. **110**(13): p. 4161-4.
105. Dai, J.H., et al., *Human Immunodeficiency Virus Integrates Directly into Naive Resting CD4(+) T Cells but Enters Naive Cells Less Efficiently than Memory Cells.* Journal of Virology, 2009. **83**(9): p. 4528-4537.
106. Colin, L. and C. Van Lint, *Molecular control of HIV-1 postintegration latency: implications for the development of new therapeutic strategies.* Retrovirology, 2009. **6**: p. 111.
107. Van Lint, C., S. Bouchat, and A. Marcello, *HIV-1 transcription and latency: an update.* Retrovirology, 2013. **10**: p. 67.
108. Mbonye, U. and J. Karn, *The Molecular Basis for Human Immunodeficiency Virus Latency.* Annu Rev Virol, 2017. **4**(1): p. 261-285.
109. Verdin, E., P. Paras, Jr., and C. Van Lint, *Chromatin disruption in the promoter of human immunodeficiency virus type 1 during transcriptional activation.* EMBO J, 1993. **12**(8): p. 3249-59.
110. Huang, J., et al., *Cellular microRNAs contribute to HIV-1 latency in resting primary CD4+ T lymphocytes.* Nat Med, 2007. **13**(10): p. 1241-7.
111. Chable-Bessia, C., et al., *Suppression of HIV-1 replication by microRNA effectors.* Retrovirology, 2009. **6**: p. 26.
112. Omoto, S. and Y.R. Fujii, *Regulation of human immunodeficiency virus 1 transcription by nef microRNA.* J Gen Virol, 2005. **86**(Pt 3): p. 751-5.
113. Hocqueloux, L., et al., *Long-term antiretroviral therapy initiated during primary HIV-1 infection is key to achieving both low HIV reservoirs and normal T cell counts.* J Antimicrob Chemother, 2013. **68**(5): p. 1169-78.
114. Ho, Y.C., et al., *Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure.* Cell, 2013. **155**(3): p. 540-51.
115. Avettand-Fenoel, V., et al., *Total HIV-1 DNA, a Marker of Viral Reservoir Dynamics with Clinical Implications.* Clin Microbiol Rev, 2016. **29**(4): p. 859-80.
116. Laird, G.M., et al., *Rapid quantification of the latent reservoir for HIV-1 using a viral outgrowth assay.* PLoS Pathog, 2013. **9**(5): p. e1003398.
117. Fun, A., et al., *A highly reproducible quantitative viral outgrowth assay for the measurement of the replication-competent latent HIV-1 reservoir.* Sci Rep, 2017. **7**: p. 43231.
118. Procopio, F.A., et al., *A Novel Assay to Measure the Magnitude of the Inducible Viral Reservoir in HIV-infected Individuals.* EBioMedicine, 2015. **2**(8): p. 874-83.
119. Cabrera, C., et al., *Rapid, Fully Automated Digital Immunoassay for p24 Protein with the Sensitivity of Nucleic Acid Amplification for Detecting Acute HIV Infection.* Clin Chem, 2015. **61**(11): p. 1372-80.
120. Norton, N.J., et al., *Innovations in the quantitative virus outgrowth assay and its use in clinical trials.* Retrovirology, 2017. **14**(1): p. 58.
121. Liszewski, M.K., J.J. Yu, and U. O'Doherty, *Detecting HIV-1 integration by repetitive-sampling Alu-gag PCR.* Methods, 2009. **47**(4): p. 254-60.
122. Eriksson, S., et al., *Comparative analysis of measures of viral reservoirs in HIV-1 eradication studies.* PLoS Pathog, 2013. **9**(2): p. e1003174.
123. Barton, K.M. and S.E. Palmer, *How to Define the Latent Reservoir: Tools of the Trade.* Curr HIV/AIDS Rep, 2016. **13**(2): p. 77-84.
124. Durand, C.M., J.N. Blankson, and R.F. Siliciano, *Developing strategies for HIV-1 eradication.* Trends Immunol, 2012. **33**(11): p. 554-62.
125. Kim, Y., J.L. Anderson, and S.R. Lewin, *Getting the "Kill" into "Shock and Kill": Strategies to Eliminate Latent HIV.* Cell Host Microbe, 2018. **23**(1): p. 14-26.

126. Archin, N.M., et al., *Administration of vorinostat disrupts HIV-1 latency in patients on antiretroviral therapy.* Nature, 2012. **487**(7408): p. 482-5.
127. Elliott, J.H., et al., *Activation of HIV transcription with short-course vorinostat in HIV-infected patients on suppressive antiretroviral therapy.* PLoS Pathog, 2014. **10**(10): p. e1004473.
128. Rasmussen, T.A., et al., *Panobinostat, a histone deacetylase inhibitor, for latent-virus reactivation in HIV-infected patients on suppressive antiretroviral therapy: a phase 1/2, single group, clinical trial.* The Lancet HIV, 2014. **1**(1): p. e13-e21.
129. Elliott, J.H., et al., *Short-term administration of disulfiram for reversal of latent HIV infection: a phase 2 dose-escalation study.* Lancet HIV, 2015. **2**(12): p. e520-9.
130. Spivak, A.M., et al., *A pilot study assessing the safety and latency-reversing activity of disulfiram in HIV-1-infected adults on antiretroviral therapy.* Clin Infect Dis, 2014. **58**(6): p. 883-90.
131. Vibholm, L., et al., *Short-Course Toll-Like Receptor 9 Agonist Treatment Impacts Innate Immunity and Plasma Viremia in Individuals With Human Immunodeficiency Virus Infection.* Clin Infect Dis, 2017. **64**(12): p. 1686-1695.
132. Bartlett, J.A., et al., *Addition of cyclophosphamide to antiretroviral therapy does not diminish the cellular reservoir in HIV-infected persons.* AIDS Res Hum Retroviruses, 2002. **18**(8): p. 535-43.
133. Vandergeeten, C., et al., *Interleukin-7 promotes HIV persistence during antiretroviral therapy.* Blood, 2013. **121**(21): p. 4321-9.
134. Geeraert, L., G. Kraus, and R.J. Pomerantz, *Hide-and-seek: the challenge of viral persistence in HIV-1 infection.* Annu Rev Med, 2008. **59**: p. 487-501.
135. Lehrman, G., et al., *Depletion of latent HIV-1 infection in vivo: a proof-of-concept study.* Lancet, 2005. **366**(9485): p. 549-55.
136. Archin, N.M., et al., *Valproic acid without intensified antiviral therapy has limited impact on persistent HIV infection of resting CD4+ T cells.* AIDS, 2008. **22**(10): p. 1131-5.
137. Siliciano, J.D., et al., *Stability of the latent reservoir for HIV-1 in patients receiving valproic acid.* J Infect Dis, 2007. **195**(6): p. 833-6.
138. Steel, A., et al., *No change to HIV-1 latency with valproate therapy.* AIDS, 2006. **20**(12): p. 1681-2.
139. Sagot-Lerolle, N., et al., *Prolonged valproic acid treatment does not reduce the size of latent HIV reservoir.* AIDS, 2008. **22**(10): p. 1125-9.
140. Shan, L., et al., *Stimulation of HIV-1-specific cytolytic T lymphocytes facilitates elimination of latent viral reservoir after virus reactivation.* Immunity, 2012. **36**(3): p. 491-501.
141. Knights, H.D.J., *A Critical Review of the Evidence Concerning the HIV Latency Reversing Effect of Disulfiram, the Possible Explanations for Its Inability to Reduce the Size of the Latent Reservoir In Vivo, and the Caveats Associated with Its Use in Practice.* AIDS Res Treat, 2017. **2017**: p. 8239428.
142. Delagreverie, H.M., et al., *Ongoing Clinical Trials of Human Immunodeficiency Virus Latency-Reversing and Immunomodulatory Agents.* Open Forum Infect Dis, 2016. **3**(4): p. ofw189.
143. Wykes, M.N. and S.R. Lewin, *Immune checkpoint blockade in infectious diseases.* Nat Rev Immunol, 2018. **18**(2): p. 91-104.
144. Laird, G.M., et al., *Ex vivo analysis identifies effective HIV-1 latency-reversing drug combinations.* J Clin Invest, 2015. **125**(5): p. 1901-12.
145. Spina, C.A., et al., *An in-depth comparison of latent HIV-1 reactivation in multiple cell model systems and resting CD4+ T cells from aviremic patients.* PLoS Pathog, 2013. **9**(12): p. e1003834.
146. Bosque, A. and V. Planelles, *Induction of HIV-1 latency and reactivation in primary memory CD4+ T cells.* Blood, 2009. **113**(1): p. 58-65.
147. Marini, A., J.M. Harper, and F. Romerio, *An in vitro system to model the establishment and reactivation of HIV-1 latency.* J Immunol, 2008. **181**(11): p. 7713-20.
148. Yang, H.C., et al., *Small-molecule screening using a human primary cell model of HIV latency identifies compounds that reverse latency without cellular activation.* Journal of Clinical Investigation, 2009. **119**(11): p. 3473-3486.

149. Tyagi, M., R.J. Pearson, and J. Karn, *Establishment of HIV latency in primary CD4+ cells is due to epigenetic transcriptional silencing and P-TEFb restriction.* J Virol, 2010. **84**(13): p. 6425-37.
150. Kim, M., et al., *A primary CD4(+) T cell model of HIV-1 latency established after activation through the T cell receptor and subsequent return to quiescence.* Nat Protoc, 2014. **9**(12): p. 2755-70.
151. Hakre, S., et al., *HIV latency: experimental systems and molecular models.* FEMS Microbiol Rev, 2012. **36**(3): p. 706-16.
152. Tsunetsugu-Yokota, Y., et al., *Homeostatically Maintained Resting Naive CD4(+) T Cells Resist Latent HIV Reactivation.* Front Microbiol, 2016. **7**: p. 1944.
153. Sahu, G.K., et al., *A novel in vitro system to generate and study latently HIV-infected long-lived normal CD4+ T-lymphocytes.* Virology, 2006. **355**(2): p. 127-37.
154. Lassen, K.G., et al., *A flexible model of HIV-1 latency permitting evaluation of many primary CD4 T-cell reservoirs.* PLoS One, 2012. **7**(1): p. e30176.
155. Shultz, L.D., et al., *Human lymphoid and myeloid cell development in NOD/LtSz-scid IL2R gamma null mice engrafted with mobilized human hemopoietic stem cells.* J Immunol, 2005. **174**(10): p. 6477-89.
156. Ishikawa, F., et al., *Development of functional human blood and immune systems in NOD/SCID/IL2 receptor {gamma} chain(null) mice.* Blood, 2005. **106**(5): p. 1565-73.
157. Audige, A., et al., *Long-term leukocyte reconstitution in NSG mice transplanted with human cord blood hematopoietic stem and progenitor cells.* BMC Immunol, 2017. **18**(1): p. 28.
158. Traggiai, E., et al., *Development of a human adaptive immune system in cord blood cell-transplanted mice.* Science, 2004. **304**(5667): p. 104-7.
159. Ito, R., et al., *Current advances in humanized mouse models.* Cell Mol Immunol, 2012. **9**(3): p. 208-14.
160. Nischang, M., et al., *Modeling HIV infection and therapies in humanized mice.* Swiss Med Wkly, 2012. **142**: p. w13618.
161. Victor Garcia, J., *Humanized mice for HIV and AIDS research.* Curr Opin Virol, 2016. **19**: p. 56-64.
162. Watanabe, S., et al., *Humanized NOD/SCID/IL2Rgamma(null) mice transplanted with hematopoietic stem cells under nonmyeloablative conditions show prolonged life spans and allow detailed analysis of human immunodeficiency virus type 1 pathogenesis.* J Virol, 2007. **81**(23): p. 13259-64.
163. Nischang, M., et al., *Humanized mice recapitulate key features of HIV-1 infection: a novel concept using long-acting anti-retroviral drugs for treating HIV-1.* PLoS One, 2012. **7**(6): p. e38853.
164. Brainard, D.M., et al., *Induction of robust cellular and humoral virus-specific adaptive immune responses in human immunodeficiency virus-infected humanized BLT mice.* J Virol, 2009. **83**(14): p. 7305-21.
165. Watanabe, S., et al., *Hematopoietic stem cell-engrafted NOD/SCID/IL2Rgamma null mice develop human lymphoid systems and induce long-lasting HIV-1 infection with specific humoral immune responses.* Blood, 2007. **109**(1): p. 212-8.
166. Sato, K., et al., *Remarkable lethal G-to-A mutations in vif-proficient HIV-1 provirus by individual APOBEC3 proteins in humanized mice.* J Virol, 2010. **84**(18): p. 9546-56.
167. Sato, K., et al., *Dynamics of memory and naive CD8+ T lymphocytes in humanized NOD/SCID/IL-2Rgammanull mice infected with CCR5-tropic HIV-1.* Vaccine, 2010. **28 Suppl 2**: p. B32-7.
168. Denton, P.W., et al., *Generation of HIV latency in humanized BLT mice.* J Virol, 2012. **86**(1): p. 630-4.
169. Marsden, M.D., et al., *HIV latency in the humanized BLT mouse.* J Virol, 2012. **86**(1): p. 339-47.
170. Choudhary, S.K., et al., *Latent HIV-1 infection of resting CD4(+) T cells in the humanized Rag2(-)/(-) gammac(-)/(-) mouse.* J Virol, 2012. **86**(1): p. 114-20.
171. Honeycutt, J.B., et al., *HIV persistence in tissue macrophages of humanized myeloid-only mice during antiretroviral therapy.* Nat Med, 2017. **23**(5): p. 638-643.
172. Villaudy, J., et al., *Long-term monitoring of the effects of HIV infection and treatment in humanized mice.* Experimental Gerontology, 2017. **94**: p. 126-126.

173. Maartens, G., C. Celum, and S.R. Lewin, *HIV infection: epidemiology, pathogenesis, treatment, and prevention.* Lancet, 2014. **384**(9939): p. 258-71.
174. Abram, M.E., et al., *Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication.* J Virol, 2010. **84**(19): p. 9864-78.
175. Mansky, L.M., *Forward mutation rate of human immunodeficiency virus type 1 in a T lymphoid cell line.* AIDS Res Hum Retroviruses, 1996. **12**(4): p. 307-14.
176. Delviks-Frankenberry, K.A., et al., *Minimal Contribution of APOBEC3-Induced G-to-A Hypermutation to HIV-1 Recombination and Genetic Variation.* PLoS Pathog, 2016. **12**(5): p. e1005646.
177. Troyer, R.M., et al., *Changes in human immunodeficiency virus type 1 fitness and genetic diversity during disease progression.* J Virol, 2005. **79**(14): p. 9006-18.
178. Richman, D.D., et al., *Rapid evolution of the neutralizing antibody response to HIV type 1 infection.* Proc Natl Acad Sci U S A, 2003. **100**(7): p. 4144-9.
179. Mulder, L.C., A. Harari, and V. Simon, *Cytidine deamination induced HIV-1 drug resistance.* Proc Natl Acad Sci U S A, 2008. **105**(14): p. 5501-6.
180. Bieniasz, P.D., *Intrinsic immunity: a front-line defense against viral attack.* Nat Immunol, 2004. **5**(11): p. 1109-15.
181. Cullen, B.R., *Role and mechanism of action of the APOBEC3 family of antiretroviral resistance factors.* J Virol, 2006. **80**(3): p. 1067-76.
182. Hache, G., L.M. Mansky, and R.S. Harris, *Human APOBEC3 proteins, retrovirus restriction, and HIV drug resistance.* AIDS Rev, 2006. **8**(3): p. 148-57.
183. Desimmie, B.A., et al., *Multiple APOBEC3 restriction factors for HIV-1 and one Vif to rule them all.* J Mol Biol, 2014. **426**(6): p. 1220-45.
184. Mangeat, B., et al., *Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts.* Nature, 2003. **424**(6944): p. 99-103.
185. Marin, M., et al., *HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation.* Nat Med, 2003. **9**(11): p. 1398-403.
186. Mehle, A., et al., *Vif overcomes the innate antiviral activity of APOBEC3G by promoting its degradation in the ubiquitin-proteasome pathway.* J Biol Chem, 2004. **279**(9): p. 7792-8.
187. Janini, M., et al., *Human Immunodeficiency Virus Type 1 DNA Sequences Genetically Damaged by Hypermutation Are Often Abundant in Patient Peripheral Blood Mononuclear Cells and May Be Generated during Near-Simultaneous Infection and Activation of CD4+ T Cells.* Journal of Virology, 2001. **75**(17): p. 7973-7986.
188. Simon, V., et al., *Natural variation in Vif: differential impact on APOBEC3G/3F and a potential role in HIV-1 diversification.* PLoS Pathog, 2005. **1**(1): p. e6.
189. Fourati, S., et al., *Partially active HIV-1 Vif alleles facilitate viral escape from specific antiretrovirals.* AIDS, 2010. **24**(15): p. 2313-21.
190. Neogi, U., et al., *Human APOBEC3G-mediated hypermutation is associated with antiretroviral therapy failure in HIV-1 subtype C-infected individuals.* J Int AIDS Soc, 2013. **16**: p. 18472.
191. Jern, P., et al., *Likely role of APOBEC3G-mediated G-to-A mutations in HIV-1 evolution and drug resistance.* PLoS Pathog, 2009. **5**(4): p. e1000367.
192. Sato, K., et al., *APOBEC3D and APOBEC3F potently promote HIV-1 diversification and evolution in humanized mouse model.* PLoS Pathog, 2014. **10**(10): p. e1004453.
193. Armitage, A.E., et al., *APOBEC3G-induced hypermutation of human immunodeficiency virus type-1 is typically a discrete "all or nothing" phenomenon.* PLoS Genet, 2012. **8**(3): p. e1002550.
194. Armitage, A.E., et al., *Possible footprints of APOBEC3F and/or other APOBEC3 deaminases, but not APOBEC3G, on HIV-1 from patients with acute/early and chronic infections.* J Virol, 2014. **88**(21): p. 12882-94.
195. Han, Y., et al., *Experimental approaches to the study of HIV-1 latency.* Nat Rev Microbiol, 2007. **5**(2): p. 95-106.
196. Hiener, B., et al., *Identification of Genetically Intact HIV-1 Proviruses in Specific CD4(+) T Cells from Effectively Treated Participants.* Cell Rep, 2017. **21**(3): p. 813-822.
197. Pallikkuth, S., et al., *Peripheral T Follicular Helper Cells Are the Major HIV Reservoir Within Central Memory CD4 T Cells in Peripheral Blood from chronic HIV infected individuals on cART.* J Virol, 2015.

198. Banga, R., et al., *PD-1(+) and follicular helper T cells are responsible for persistent HIV-1 transcription in treated aviremic individuals.* Nat Med, 2016. **22**(7): p. 754-61.
199. Zhang, J. and A.S. Perelson, *Contribution of follicular dendritic cells to persistent HIV viremia.* J Virol, 2013. **87**(14): p. 7893-901.
200. McNamara, L.A., et al., *CD133+ hematopoietic progenitor cells harbor HIV genomes in a subset of optimally treated people with long-term viral suppression.* J Infect Dis, 2013. **207**(12): p. 1807-16.
201. Yukl, S.A., et al., *A comparison of methods for measuring rectal HIV levels suggests that HIV DNA resides in cells other than CD4+ T cells, including myeloid cells.* AIDS, 2014. **28**(3): p. 439-42.
202. Zalar, A., et al., *Macrophage HIV-1 infection in duodenal tissue of patients on long term HAART.* Antiviral Res, 2010. **87**(2): p. 269-71.
203. Chun, T.W., et al., *Persistence of HIV in gut-associated lymphoid tissue despite long-term antiretroviral therapy.* J Infect Dis, 2008. **197**(5): p. 714-20.
204. Lusic, M. and R.F. Siliciano, *Nuclear landscape of HIV-1 infection and integration.* Nat Rev Microbiol, 2017. **15**(2): p. 69-82.
205. Kumar, N., A. Chahroudi, and G. Silvestri, *Animal models to achieve an HIV cure.* Curr Opin HIV AIDS, 2016. **11**(4): p. 432-41.
206. Evans, D.T. and G. Silvestri, *Nonhuman primate models in AIDS research.* Curr Opin HIV AIDS, 2013. **8**(4): p. 255-61.
207. Ito, M., et al., *NOD/SCID/gamma(c)(null) mouse: an excellent recipient mouse model for engraftment of human cells.* Blood, 2002. **100**(9): p. 3175-82.
208. Myburgh, R., et al., *Lentivector Knockdown of CCR5 in Hematopoietic Stem and Progenitor Cells Confers Functional and Persistent HIV-1 Resistance in Humanized Mice.* J Virol, 2015. **89**(13): p. 6761-72.
209. Sauer, B., *Inducible gene targeting in mice using the Cre/lox system.* Methods-a Companion to Methods in Enzymology, 1998. **14**(4): p. 381-392.
210. Sauer, B., *Cre/lox: one more step in the taming of the genome.* Endocrine, 2002. **19**(3): p. 221-8.
211. Missirlis, P.I., D.E. Smailus, and R.A. Holt, *A high-throughput screen identifying sequence and promiscuity characteristics of the loxP spacer region in Cre-mediated recombination.* BMC Genomics, 2006. **7**: p. 73.
212. Sharaf, R., T.R. Mempel, and T.T. Murooka, *Visualizing the Behavior of HIV-Infected T Cells In Vivo Using Multiphoton Intravital Microscopy.* Methods Mol Biol, 2016. **1354**: p. 189-201.
213. Mattapallil, J.J., et al., *Massive infection and loss of memory CD4+ T cells in multiple tissues during acute SIV infection.* Nature, 2005. **434**(7037): p. 1093-7.
214. Simmonds, P., et al., *Human Immunodeficiency Virus-Infected Individuals Contain Provirus in Small Numbers of Peripheral Mononuclear-Cells and at Low Copy Numbers.* Journal of Virology, 1990. **64**(2): p. 864-872.
215. Barde, I., P. Salmon, and D. Trono, *Production and titration of lentiviral vectors.* Curr Protoc Neurosci, 2010. **Chapter 4**: p. Unit 4 21.
216. Grandadam, M., et al., *Dose-dependent systemic human immunodeficiency virus infection of SCID-hu mice after intraperitoneal virus injection.* Research in Virology, 1995. **146**(2): p. 101-112.
217. Sallusto, F., et al., *Two subsets of memory T lymphocytes with distinct homing potentials and effector functions.* Nature, 1999. **401**(6754): p. 708-12.
218. Haqqani, A.A., et al., *Central memory CD4+ T cells are preferential targets of double infection by HIV-1.* Virol J, 2015. **12**: p. 184.
219. Douek, D.C., et al., *HIV preferentially infects HIV-specific CD4+ T cells.* Nature, 2002. **417**(6884): p. 95-8.
220. Hockett, R.D., et al., *Constant mean viral copy number per infected cell in tissues regardless of high, low, or undetectable plasma HIV RNA.* J Exp Med, 1999. **189**(10): p. 1545-54.
221. Brenchley, J.M., et al., *T-Cell Subsets That Harbor Human Immunodeficiency Virus (HIV) In Vivo: Implications for HIV Pathogenesis.* Journal of Virology, 2004. **78**(3): p. 1160-1168.
222. *Whole Genome Amplification*, in *Methods and Protocols*, T. Kroneis, Editor.

223. Hodel, F., et al., *HIV-1 latent reservoir: size matters.* Future Virol, 2016. **11**(12): p. 785-794.
224. Passaes, C.P., et al., *Ultrasensitive HIV-1 p24 Assay Detects Single Infected Cells and Differences in Reservoir Induction by Latency Reversal Agents.* J Virol, 2017. **91**(6).
225. Miles, B., S.M. Miller, and E. Connick, *CD4 T Follicular Helper and Regulatory Cell Dynamics and Function in HIV Infection.* Front Immunol, 2016. **7**: p. 659.
226. Colineau, L., et al., *HIV-Infected Spleens Present Altered Follicular Helper T Cell (Tfh) Subsets and Skewed B Cell Maturation.* PLoS One, 2015. **10**(10): p. e0140978.
227. Chevalier, M.F. and L. Weiss, *The split personality of regulatory T cells in HIV infection.* Blood, 2013. **121**(1): p. 29-37.
228. Hataye, J., *HIV Exponential Growth Dependent on Burst Size Breakthrough of the Allee Threshold*, in *Strategies for an HIV Cure*. 2016, NIH National Institute of Allergy and Infectious Diseases.
229. Wu, Y. and J.W. Marsh, *Gene transcription in HIV infection.* Microbes and Infection, 2003. **5**(11): p. 1023-1027.
230. Wu, Y. and J.W. Marsh, *Selective transcription and modulation of resting T cell activity by preintegrated HIV DNA.* Science, 2001. **293**(5534): p. 1503-6.
231. Schnittman, S.M., et al., *Preferential infection of CD4+ memory T cells by human immunodeficiency virus type 1: evidence for a role in the selective T-cell functional defects observed in infected individuals.* Proc Natl Acad Sci U S A, 1990. **87**(16): p. 6058-62.
232. Jones, S., et al., *Lentiviral vector design for optimal T cell receptor gene expression in the transduction of peripheral blood lymphocytes and tumor-infiltrating lymphocytes.* Hum Gene Ther, 2009. **20**(6): p. 630-40.
233. Salmon, P., et al., *High-level transgene expression in human hematopoietic progenitors and differentiated blood lineages after transduction with improved lentiviral vectors.* Blood, 2000. **96**(10): p. 3392-8.
234. Gilham, D.E., et al., *Cytokine stimulation and the choice of promoter are critical factors for the efficient transduction of mouse T cells with HIV-1 vectors.* J Gene Med, 2010. **12**(2): p. 129-36.
235. Morrissey, D., et al., *Plasmid Transgene Expression in vivo: Promoter and Tissue Variables.* 2013.
236. Kimura, T., et al., *Lentiviral vectors with CMV or MHCII promoters administered in vivo: immune reactivity versus persistence of expression.* Mol Ther, 2007. **15**(7): p. 1390-9.
237. Brooks, A.R., et al., *Transcriptional silencing is associated with extensive methylation of the CMV promoter following adenoviral gene delivery to muscle.* J Gene Med, 2004. **6**(4): p. 395-404.
238. Chang, L., et al., *Simple diffusion-constrained immunoassay for p24 protein with the sensitivity of nucleic acid amplification for detecting acute HIV infection.* J Virol Methods, 2013. **188**(1-2): p. 153-60.
239. Fischer, M., et al., *Cellular viral rebound after cessation of potent antiretroviral therapy predicted by levels of multiply spliced HIV-1 RNA encoding nef.* J Infect Dis, 2004. **190**(11): p. 1979-88.
240. Dobrowsky, T.M., et al., *Adhesion and fusion efficiencies of human immunodeficiency virus type 1 (HIV-1) surface proteins.* Sci Rep, 2013. **3**: p. 3014.
241. Bosque, A., et al., *Benzotriazoles Reactivate Latent HIV-1 through Inactivation of STAT5 SUMOylation.* Cell Rep, 2017. **18**(5): p. 1324-1334.
242. Rochat, M.A., E. Schlaepfer, and R.F. Speck, *Promising Role of Toll-Like Receptor 8 Agonist in Concert with Prostratin for Activation of Silent HIV.* J Virol, 2017. **91**(4).
243. Noel, N., et al., *Long-Term Spontaneous Control of HIV-1 Is Related to Low Frequency of Infected Cells and Inefficient Viral Reactivation.* J Virol, 2016. **90**(13): p. 6148-58.
244. Li, Q., et al., *Comparison of the sorting efficiency and influence on cell function between the sterile flow cytometry and immunomagnetic bead purification methods.* Prep Biochem Biotechnol, 2013. **43**(2): p. 197-206.
245. Winnard, P.T., et al., *Development of novel chimeric transmembrane proteins for multimodality imaging of cancer cells.* Cancer Biology & Therapy, 2014. **6**(12): p. 1889-1899.
246. Wang, X., et al., *A transgene-encoded cell surface polypeptide for selection, in vivo tracking, and ablation of engineered cells.* Blood, 2011. **118**(5): p. 1255-63.

247. Cranfill, P.J., et al., *Quantitative assessment of fluorescent proteins.* Nat Methods, 2016. **13**(7): p. 557-62.
248. Jaisser, F., *Inducible gene expression and gene modification in transgenic mice.* J Am Soc Nephrol, 2000. **11 Suppl 16**: p. S95-S100.
249. Dragatsis, I. and S. Zeitlin, *A method for the generation of conditional gene repair mutations in mice.* Nucleic Acids Res, 2001. **29**(3): p. E10.
250. Oberdoerffer, P., *Unidirectional Cre-mediated genetic inversion in mice using the mutant loxP pair lox66/lox71.* Nucleic Acids Research, 2003. **31**(22): p. 140e-140.
251. Dabrowski, M., Z. Bukowy-Bieryllo, and E. Zietkiewicz, *Translational readthrough potential of natural termination codons in eucaryotes--The impact of RNA sequence.* RNA Biol, 2015. **12**(9): p. 950-8.
252. Kaczmarczyk, S.J. and J.E. Green, *A single vector containing modified cre recombinase and LOX recombination sequences for inducible tissue-specific amplification of gene expression.* Nucleic Acids Res, 2001. **29**(12): p. E56-6.
253. Lakso, M., et al., *Targeted oncogene activation by site-specific recombination in transgenic mice.* Proc Natl Acad Sci U S A, 1992. **89**(14): p. 6232-6.
254. Karn, J. and C.M. Stoltzfus, *Transcriptional and posttranscriptional regulation of HIV-1 gene expression.* Cold Spring Harb Perspect Med, 2012. **2**(2): p. a006916.
255. McDougal, J.S., et al., *Immunoassay for the detection and quantitation of infectious human retrovirus, lymphadenopathy-associated virus (LAV).* J Immunol Methods, 1985. **76**(1): p. 171-83.
256. Phuong, T. *Understand TCID50 and MOI*. I am a novice 2011 [cited 2015; Available from: https://iamanovice.wordpress.com/2011/03/31/tcid50/.
257. Kaiser, P., et al., *Productive human immunodeficiency virus type 1 infection in peripheral blood predominantly takes place in CD4/CD8 double-negative T lymphocytes.* J Virol, 2007. **81**(18): p. 9693-706.
258. Christopherson, C., et al., *PCR-Based assay to quantify human immunodeficiency virus type 1 DNA in peripheral blood mononuclear cells.* J Clin Microbiol, 2000. **38**(2): p. 630-4.
259. Althaus, C.F., et al., *Rational design of HIV-1 fluorescent hydrolysis probes considering phylogenetic variation and probe performance.* J Virol Methods, 2010. **165**(2): p. 151-60.
260. Laird, G.M., et al., *Measuring the Frequency of Latent HIV-1 in Resting CD4(+) T Cells Using a Limiting Dilution Coculture Assay.* Methods Mol Biol, 2016. **1354**: p. 239-53.
261. van Zyl, G., M.J. Bale, and M.F. Kearney, *HIV evolution and diversity in ART-treated patients.* Retrovirology, 2018. **15**(1): p. 14.
262. Krisko, J.F., et al., *APOBEC3G and APOBEC3F Act in Concert To Extinguish HIV-1 Replication.* J Virol, 2016. **90**(9): p. 4681-95.
263. Krisko, J.F., et al., *HIV restriction by APOBEC3 in humanized mice.* PLoS Pathog, 2013. **9**(3): p. e1003242.
264. Albin, J.S. and R.S. Harris, *Interactions of host APOBEC3 restriction factors with HIV-1 in vivo: implications for therapeutics.* Expert Rev Mol Med, 2010. **12**: p. e4.
265. Kim, E.Y., et al., *Human APOBEC3 induced mutation of human immunodeficiency virus type-1 contributes to adaptation and evolution in natural infection.* PLoS Pathog, 2014. **10**(7): p. e1004281.
266. Alteri, C., et al., *Incomplete APOBEC3G/F Neutralization by HIV-1 Vif Mutants Facilitates the Genetic Evolution from CCR5 to CXCR4 Usage.* Antimicrob Agents Chemother, 2015. **59**(8): p. 4870-81.
267. Sadler, H.A., et al., *APOBEC3G contributes to HIV-1 variation through sublethal mutagenesis.* J Virol, 2010. **84**(14): p. 7396-404.
268. Kim, E.Y., et al., *Human APOBEC3G-mediated editing can promote HIV-1 sequence diversification and accelerate adaptation to selective pressure.* J Virol, 2010. **84**(19): p. 10402-5.
269. Russell, R.A., et al., *APOBEC3G induces a hypermutation gradient: purifying selection at multiple steps during HIV-1 replication results in levels of G-to-A mutations that are high in DNA, intermediate in cellular viral RNA, and low in virion RNA.* Retrovirology, 2009. **6**: p. 16.
270. Berkhout, B. and A. de Ronde, *APOBEC3G versus reverse transcriptase in the generation of HIV-1 drug-resistance mutations.* AIDS, 2004. **18**(13): p. 1861-3.

271. Lecossier, D., et al., *Hypermutation of HIV-1 DNA in the absence of the Vif protein.* Science, 2003. **300**(5622): p. 1112.
272. Martinez, M.A., J.P. Vartanian, and S. Wainhobson, *Hypermutagenesis of Rna Using Human-Immunodeficiency-Virus Type-1 Reverse-Transcriptase and Biased Dntp Concentrations.* Proceedings of the National Academy of Sciences of the United States of America, 1994. **91**(25): p. 11787-11791.
273. Keulen, W., et al., *Initial appearance of the 184Ile variant in lamivudine-treated patients is caused by the mutational bias of human immunodeficiency virus type 1 reverse transcriptase.* Journal of Virology, 1997. **71**(4): p. 3346-3350.
274. Vartanian, J.P., et al., *Selection, recombination, and G----A hypermutation of human immunodeficiency virus type 1 genomes.* J Virol, 1991. **65**(4): p. 1779-88.
275. Schinazi, R.F., et al., *Characterization of human immunodeficiency viruses resistant to oxathiolane-cytosine nucleosides.* Antimicrob Agents Chemother, 1993. **37**(4): p. 875-81.
276. Schinazi, R.F., et al., *Synthesis and virucidal activity of a water-soluble, configurationally stable, derivatized C60 fullerene.* Antimicrob Agents Chemother, 1993. **37**(8): p. 1707-10.
277. Schuurman, R., et al., *Rapid changes in human immunodeficiency virus type 1 RNA load and appearance of drug-resistant virus populations in persons treated with lamivudine (3TC).* J Infect Dis, 1995. **171**(6): p. 1411-9.
278. Frost, S.D., et al., *Evolution of lamivudine resistance in human immunodeficiency virus type 1-infected individuals: the relative roles of drift and selection.* J Virol, 2000. **74**(14): p. 6262-8.
279. Rongvaux, A., et al., *Human hemato-lymphoid system mice: current use and future potential for medicine.* Annu Rev Immunol, 2013. **31**: p. 635-674.
280. Henriet, S., et al., *Tumultuous relationship between the human immunodeficiency virus type 1 viral infectivity factor (Vif) and the human APOBEC-3G and APOBEC-3F restriction factors.* Microbiol Mol Biol Rev, 2009. **73**(2): p. 211-32.
281. Feng, Y., et al., *Suppression of APOBEC3-mediated restriction of HIV-1 by Vif.* Front Microbiol, 2014. **5**: p. 450.
282. Russell, R.A. and V.K. Pathak, *Identification of two distinct human immunodeficiency virus type 1 Vif determinants critical for interactions with human APOBEC3G and APOBEC3F.* J Virol, 2007. **81**(15): p. 8201-10.
283. Binka, M., et al., *The activity spectrum of Vif from multiple HIV-1 subtypes against APOBEC3G, APOBEC3F, and APOBEC3H.* J Virol, 2012. **86**(1): p. 49-59.
284. Ooms, M., et al., *HIV-1 Vif adaptation to human APOBEC3H haplotypes.* Cell Host Microbe, 2013. **14**(4): p. 411-21.
285. Ooms, M., et al., *The resistance of human APOBEC3H to HIV-1 NL4-3 molecular clone is determined by a single amino acid in Vif.* PLoS One, 2013. **8**(2): p. e57744.
286. OhAinle, M., et al., *Antiretroelement activity of APOBEC3H was lost twice in recent human evolution.* Cell Host Microbe, 2008. **4**(3): p. 249-59.
287. Wang, X., et al., *Analysis of human APOBEC3H haplotypes and anti-human immunodeficiency virus type 1 activity.* J Virol, 2011. **85**(7): p. 3142-52.
288. Jabara, C.B., et al., *Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID.* Proc Natl Acad Sci U S A, 2011. **108**(50): p. 20166-71.
289. Nei, M. and W.H. Li, *Mathematical model for studying genetic variation in terms of restriction endonucleases.* Proceedings of the National Academy of Sciences, 1979. **76**(10): p. 5269-5273.
290. Refsland, E.W., J.F. Hultquist, and R.S. Harris, *Endogenous origins of HIV-1 G-to-A hypermutation and restriction in the nonpermissive T cell line CEM2n.* PLoS Pathog, 2012. **8**(7): p. e1002800.
291. Bednar, M.M., et al., *Compartmentalization, Viral Evolution, and Viral Latency of HIV in the CNS.* Curr HIV/AIDS Rep, 2015. **12**(2): p. 262-71.
292. Cuevas, J.M., et al., *Extremely High Mutation Rate of HIV-1 In Vivo.* PLoS Biol, 2015. **13**(9): p. e1002251.
293. Kieffer, T.L., et al., *G-->A hypermutation in protease and reverse transcriptase regions of human immunodeficiency virus type 1 residing in resting CD4+ T cells in vivo.* J Virol, 2005. **79**(3): p. 1975-80.

294. Imamichi, H., et al., *Defective HIV-1 proviruses produce novel protein-coding RNA species in HIV-infected patients on combination antiretroviral therapy.* Proc Natl Acad Sci U S A, 2016. **113**(31): p. 8783-8.
295. Herbeuval, J.P., et al., *CD4+ T-cell death induced by infectious and noninfectious HIV-1: role of type 1 interferon-dependent, TRAIL/DR5-mediated apoptosis.* Blood, 2005. **106**(10): p. 3524-31.
296. Finzi, D., S.F. Plaeger, and C.W. Dieffenbach, *Defective virus drives human immunodeficiency virus infection, persistence, and pathogenesis.* Clin Vaccine Immunol, 2006. **13**(7): p. 715-21.
297. Quan, Y., et al., *Highly diversified multiply drug-resistant HIV-1 quasispecies in PBMCs: a case report.* Retrovirology, 2008. **5**: p. 43.
298. Mansky, L.M. and L.C. Bernard, *3 '-azido-3 '-deoxythymidine (AZT) and AZT-resistant reverse transcriptase can increase the in vivo mutation rate of human immunodeficiency virus type 1.* Journal of Virology, 2000. **74**(20): p. 9532-9539.
299. Josefsson, L., et al., *Single cell analysis of lymph node tissue from HIV-1 infected patients reveals that the majority of CD4+ T-cells contain one HIV-1 DNA molecule.* PLoS Pathog, 2013. **9**(6): p. e1003432.
300. Gunthard, H.F., et al., *Residual human immunodeficiency virus (HIV) Type 1 RNA and DNA in lymph nodes and HIV RNA in genital secretions and in cerebrospinal fluid after suppression of viremia for 2 years.* J Infect Dis, 2001. **183**(9): p. 1318-27.
301. Mehandru, S., et al., *Primary HIV-1 infection is associated with preferential depletion of CD4+ T lymphocytes from effector sites in the gastrointestinal tract.* J Exp Med, 2004. **200**(6): p. 761-70.
302. Sigal, A., et al., *Cell-to-cell spread of HIV permits ongoing replication despite antiretroviral therapy.* Nature, 2011. **477**(7362): p. 95-8.
303. Yu, X., et al., *Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex.* Science, 2003. **302**(5647): p. 1056-60.
304. Zhang, J., et al., *PEAR: a fast and accurate Illumina Paired-End reAd mergeR.* Bioinformatics, 2014. **30**(5): p. 614-20.
305. Librado, P. and J. Rozas, *DnaSP v5: a software for comprehensive analysis of DNA polymorphism data.* Bioinformatics, 2009. **25**(11): p. 1451-2.
306. Chun, T.W., *HIV-infected individuals receiving effective antiviral therapy for extended periods of time continually replenish their viral reservoir.* J Clin Invest, 2005. **115**(11): p. 3250-5.
307. Thome, J.J. and D.L. Farber, *Emerging concepts in tissue-resident T cells: lessons from humans.* Trends Immunol, 2015. **36**(7): p. 428-35.
308. Smith, J.L., et al., *Multiple ways of targeting APOBEC3-virion infectivity factor interactions for anti-HIV-1 drug development.* Trends Pharmacol Sci, 2009. **30**(12): p. 638-46.
309. Stopak, K.S., et al., *Distinct patterns of cytokine regulation of APOBEC3G expression and activity in primary lymphocytes, macrophages, and dendritic cells.* J Biol Chem, 2007. **282**(6): p. 3539-46.
310. Koning, F.A., et al., *Defining APOBEC3 expression patterns in human tissues and hematopoietic cell subsets.* J Virol, 2009. **83**(18): p. 9474-85.